

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 14:12:28 ; Search time 25.025 Seconds  
(without alignments)  
158.069 Million cell updates/sec

Title: US-09-745-008-14

Perfect score: 78

Sequence: 1 RQRLPKRMGGSYRC 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	14	7 ADC21513	Adc21513 T. cruzi
2	78	100.0	21	7 ADC21512	Adc21512 T. cruzi
3	78	100.0	1162	7 ADC21533	Adc21533 T. cruzi
4	67.5	86.5	45	7 ADC21511	Adc21511 T. cruzi
5	42	53.8	30	2 AAW89049	AAW89049 Polypepti
6	42	53.8	30	4 ABB51214	Abb51214 Human sec
7	42	53.8	30	6 ABO45471	Abo45471 Novel hum
8	42	53.8	30	7 ABO26951	Abo26951 Protein a
9	42	53.8	339	4 ABB68550	Abb68550 Drosophil
10	41	52.6	189	4 AAU42341	Aau42341 Propionib
11	41	52.6	189	6 ABM38860	Abm38860 Propionib
12	41	52.6	212	5 ABB75629	Abb75629 Human end
13	40.5	51.9	235	2 AAR05328	Aar05328 Rat tumou
14	40.5	51.9	235	2 AAR05996	Aar05996 Rat tumou
15	40.5	51.9	235	5 AAU10546	Aau10546 Rat tumou
16	40	51.3	92	3 AAG33234	Aag33234 Zea mays
17	40	51.3	138	3 AAG33233	Aag33233 Zea mays
18	40	51.3	151	4 ABG16755	Abg16755 Novel hum
19	40	51.3	230	4 AAG98324	Aag98324 Escherich
20	40	51.3	230	6 ABU14845	Abu14845 Protein e
21	40	51.3	259	7 ADB64149	Adb64149 Human pro
22	40	51.3	279	7 ABM74436	Abm74436 DNA clone
23	40	51.3	299	4 ABB76024	Abb76024 Human pro
24	40	51.3	302	6 ABUS2331	Abu52331 Human GPC
25	40	51.3	302	6 ABUS2330	Abu52330 Human GPC

#### ALIGNMENTS

RESULT 1

ADC21513

ID ADC21513 standard; peptide; 14 AA.

XX AC ADC21513;

DT 18-DEC-2003 (first entry)

DE T. cruzi trans-sialidase, TS, neurotrophic peptide C14.

KW Trans-sialidase; TS; neurotrophic peptide; interleukin-6;  
KW IL-6 secretion inducing peptide; neuron; glial cell; trophic support;  
KW ciliary neurotrophic factor; CNTF; leukaemia inhibitory factor; LIF;  
KW anyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;  
KW Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;  
KW multiple sclerosis; stroke; brain trauma; spinal cord trauma;  
KW peripheral nerve trauma.

OS Trypanosoma cruzi.

XX US2002137667-A1.

XX 26-SEP-2002.

XX 20-DEC-2000; 2000US-00745008. →

XX 20-DEC-1999; 99US-0172881P.

XX (TUFT ) UNIV TUFTS.

XX Chuenkova M, Pereira MA;

XX WPI; 2003-786654/74.

XX T. cruzi trans-sialidase derived neurotrophic and interleukin-6 secretion  
XX inducing peptides, useful in the treatment of neuronal degeneration  
XX caused for example by Chagas' disease.

PS Claim 5; SEQ ID NO 14; 79pp; English.

XX The invention relates to a T. cruzi trans-sialidase (TS) derived  
XX neurotrophic peptide appearing as ADC21513 or ADC21511 called C14 and  
XX C14, or its variant, and an interleukin-6 (IL-6) secretion inducing  
XX peptide appearing as ADC21513 called TR-1 (terminal repeat 1). Also  
XX included are a composition comprising the peptides, a fusion protein  
XX comprising the peptides and a fusion partner, a composition comprising  
XX the fusion protein and a physiological acceptable carrier, providing  
XX trophic support for neurons or glial cells in a mammal (comprising

Abb55482 Lactococ  
Aag43982 Zea mays  
Aag43981 Zea mays  
Aap43486 Human sec  
Ada54123 Human pro  
Aag43980 Zea mays  
Abb76023 Neurotrim  
Abb76018 Neurotrim  
Adj10922 Human cel  
Aao30844 Human cel  
Abg12100 Novel hum  
Abu52336 Human GPC  
Abu52329 Human GPC  
Aay42697 Mouse ser  
Aay42696 Rat serin  
Aab65622 Novel pro  
Aau66518 Propionib  
Aau40677 Propionib  
Abm37196 Propionib  
Abm63037 Propionib

26 40 51.3 428 5 ABB55482  
27 40 51.3 483 3 AAG43982  
28 40 51.3 529 3 AAG43981  
29 40 51.3 559 5 ABP43486  
30 40 51.3 559 6 ADA54123  
31 40 51.3 565 3 AAG43980  
32 40 51.3 570 4 ABB76023  
33 40 51.3 586 4 ABB76018  
34 40 51.3 869 5 ABJ10922  
35 40 51.3 886 7 AAO30844  
36 40 51.3 1219 4 ABG12100  
37 40 51.3 1315 6 ABUS2336  
38 40 51.3 1386 6 ABUS2329  
39 39.5 50.6 732 2 AAY42697  
40 39.5 50.6 733 2 AAY42696  
41 39.5 50.6 806 4 AAB65622  
42 39 50.0 71 4 AAU66518  
43 39 50.0 71 4 AAU40677  
44 39 50.0 71 6 ABM37196  
45 39 50.0 71 6 ABM63037

CC administering a therapeutically effective amount of T. cruzi trans-  
 CC sialidase (TS) or a neurotrophic variant) and stimulating (M3) the  
 CC secretion of IL-6 (interleukin-6) in a mammal comprising administering TS  
 CC or an IL-6 inducing variant. The fusion partner comprises a mammalian  
 CC neurotrophic factor which is ciliary neurotrophic factor (CNTF) or  
 CC leukaemia inhibitory factor (LIF). The peptides are useful in providing  
 CC trophic support for neurons and glial cells in a mammal suffering a  
 CC condition selected from: amyotrophic lateral sclerosis, Alzheimer's  
 CC disease, Parkinson's disease, Huntington's disease, Chagas' disease,  
 CC peripheral neuropathy, palsies, multiple sclerosis, stroke, brain trauma,  
 CC spinal cord trauma and peripheral nerve trauma, and in stimulating the  
 CC secretion of IL-6. The present sequence represents a trans-sialidase  
 CC neurotrophic peptide.

XX Sequence 14 AA;

Query Match 100.0%; Score 78; DB 7; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGGSYRC 14  
 |||||  
 Db 1 RQRLPKRMGGSYRC 14

# RESULT 2

ID ADC21512 standard; peptide; 21 AA.

XX

AC ADC21512;

DT 18-DEC-2003 (first entry)

DE T. cruzi trans-sialidase, TS, neurotrophic peptide CPN1.

XX Trans-sialidase; TS; neurotrophic peptide; interleukin-6;

KW IL-6 secretion inducing peptide; neuron; glial cell; trophic support;  
 KW ciliary neurotrophic factor; CNTF; leukaemia inhibitory factor; LIF;  
 KW amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;  
 KW Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;  
 KW multiple sclerosis; stroke; brain trauma; spinal cord trauma;  
 KW peripheral nerve trauma.

XX Trypanosoma cruzi.

XX US2002137667-A1.

XX 26-SEP-2002.

XX 20-DEC-2000; 2000US-00745008.

XX 20-DEC-1999; 99US-0172881P.

XX (TUFT ) UNIV TUFTS.

XX Chuenkova M, Pereira MA;

XX WPI; 2003-786654/74.

XX T. cruzi trans-sialidase derived neurotrophic and interleukin-6 secretion  
 PT inducing peptides, useful in the treatment of neuronal degeneration  
 PT caused for example by Chagas' disease.

XX Example 1; SEQ ID NO 13; 79pp; English.

XX The invention relates to a T. cruzi trans-sialidase (TS) derived  
 CC neurotrophic peptide appearing as ADC21513 or ADC21511 called C44 and  
 CC C14, or its variant, and an interleukin-6 (IL-6) secretion inducing  
 CC peptide appearing as ADC21531 called TR-1 (terminal repeat 1). Also  
 CC included are a composition comprising the peptides, fusion protein  
 CC comprising the peptides and a fusion partner, a composition comprising  
 CC the fusion protein and a physiological acceptable carrier, providing  
 CC trophic support for neurons or glial cells in a mammal (comprising

CC administering a therapeutically effective amount of T. cruzi trans-  
 CC sialidase (TS) or a neurotrophic variant) and stimulating (M3) the  
 CC secretion of IL-6 (interleukin-6) in a mammal comprising administering TS  
 CC or an IL-6 inducing variant. The fusion partner comprises a mammalian  
 CC neurotrophic factor which is ciliary neurotrophic factor (CNTF) or  
 CC leukaemia inhibitory factor (LIF). The peptides are useful in providing  
 CC trophic support for neurons and glial cells in a mammal suffering a  
 CC condition selected from: amyotrophic lateral sclerosis, Alzheimer's  
 CC disease, Parkinson's disease, Huntington's disease, Chagas' disease,  
 CC peripheral neuropathy, palsies, multiple sclerosis, stroke, brain trauma,  
 CC spinal cord trauma and peripheral nerve trauma, and in stimulating the  
 CC secretion of IL-6. The present sequence represents a trans-sialidase  
 CC neurotrophic peptide.

XX Sequence 21 AA;

Query Match 100.0%; Score 78; DB 7; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGGSYRC 14  
 |||||  
 Db 1 RQRLPKRMGGSYRC 14

# RESULT 3

ID ADC21533 standard; protein; 1162 AA.

XX

AC ADC21533;

DT 18-DEC-2003 (first entry)

DE T. cruzi trans-sialidase, TS, clone 7F.

XX Trans-sialidase; TS; neurotrophic peptide; interleukin-6;

KW IL-6 secretion inducing peptide; neuron; glial cell; trophic support;  
 KW ciliary neurotrophic factor; CNTF; leukaemia inhibitory factor; LIF;  
 KW amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;  
 KW Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;  
 KW multiple sclerosis; stroke; brain trauma; spinal cord trauma;  
 KW peripheral nerve trauma.

XX Trypanosoma cruzi.

XX US2002137667-A1.

XX 26-SEP-2002.

XX 20-DEC-2000; 2000US-00745008.

XX 20-DEC-1999; 99US-0172881P.

XX (TUFT ) UNIV TUFTS.

XX Chuenkova M, Pereira MA;

XX WPI; 2003-786654/74.

XX N-PSDB; ADC21532.

XX T. cruzi trans-sialidase derived neurotrophic and interleukin-6 secretion  
 PT inducing peptides, useful in the treatment of neuronal degeneration  
 PT caused for example by Chagas' disease.

XX Disclosure; SEQ ID NO 34; 79pp; English.

XX The invention relates to a T. cruzi trans-sialidase (TS) derived  
 CC neurotrophic peptide appearing as ADC21513 or ADC21511 called C44 and  
 CC C14, or its variant, and an interleukin-6 (IL-6) secretion inducing  
 CC peptide appearing as ADC21531 called TR-1 (terminal repeat 1). Also  
 CC included are a composition comprising the peptides, fusion protein  
 CC comprising the peptides and a fusion partner, a composition comprising  
 CC the fusion protein and a physiological acceptable carrier, providing  
 CC the fusion protein and a physiological acceptable carrier, providing

CC trophic support for neurons or glial cells in a mammal (comprising  
 CC administering a therapeutically effective amount of T. cruzi trans-  
 CC sialidase (TS) or a neurotrophic variant) and stimulating (M3) the  
 CC secretion of IL-6 (interleukin-6) in a mammal comprising administering TS  
 CC or an IL-6 inducing variant. The fusion partner comprises a mammalian  
 CC neurotrophic factor which is ciliary neurotrophic factor (CNTF) or  
 CC leukaemia inhibitory factor (LIF). The peptides are useful in providing  
 CC trophic support for neurons and glial cells in a mammal suffering a  
 CC condition selected from: amyotrophic lateral sclerosis, Alzheimer's  
 CC disease, Parkinson's disease, Huntington's disease, Chagas' disease,  
 CC peripheral neuropathy, palsies, multiple sclerosis, stroke, brain trauma,  
 CC spinal cord trauma and peripheral nerve trauma, and in stimulating the  
 CC secretion of IL-6. The present sequence represents trans-sialidase clone  
 CC 7F.  
 XX  
 SQ Sequence 1162 AA;

Query Match 100.0%; Score 78; DB 7; Length 1162;  
 Best Local Similarity 100.0%; Pred. No. 8.2e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RQRLPRMGGSYRC 14  
 |||||  
 Db 379 RQRLPRMGGSYRC 392

RESULT 4  
 ADC21511  
 ID ADC21511 standard; peptide; 45 AA.  
 XX  
 AC ADC21511;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE T. cruzi trans-sialidase, TS, neurotrophic peptide C44.  
 XX  
 KW Trans-sialidase; TS; neurotrophic peptide; interleukin-6;  
 KW IL-6 secretion inducing peptide; neuron; glial cell; trophic support;  
 KW ciliary neurotrophic factor; CNTF; leukaemia inhibitory factor; LIF;  
 KW amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;  
 KW Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;  
 KW multiple sclerosis; stroke; brain trauma; spinal cord trauma;  
 KW peripheral nerve trauma.  
 XX  
 OS Trypanosoma cruzi.  
 XX  
 PN US2002137667-A1.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 20-DEC-2000; 2000US-00745008.  
 XX  
 PR 20-DEC-1999; 99US-0172881P.  
 XX  
 PA (TUFT ) UNIV TUFTS.  
 XX  
 PI Chuenkova M, Pereira MA;  
 XX  
 XX WPI; 2003-786654/74.  
 DR  
 XX  
 PT T. cruzi trans-sialidase derived neurotrophic and interleukin-6 secretion  
 PT inducing peptides, useful in the treatment of neuronal degeneration  
 PT caused for example by Chagas' disease.  
 XX  
 PS Claim 4; SEQ ID NO 12; 79pp; English.  
 XX  
 CC The invention relates to a T. cruzi trans-sialidase (TS) derived  
 CC neurotrophic peptide appearing as ADC21513 or ADC21511 called C44 and  
 CC C14, or its variant, and an interleukin-6 (IL-6) secretion inducing  
 CC peptide appearing as ADC21531 called TR-1 (terminal repeat 1). Also  
 CC included are a composition comprising the peptides, fusion protein  
 CC comprising the peptides and a fusion partner, a composition comprising  
 CC the fusion protein and a physiological acceptable carrier, providing

CC trophic support for neurons or glial cells in a mammal (comprising  
 CC administering a therapeutically effective amount of T. cruzi trans-  
 CC sialidase (TS) or a neurotrophic variant) and stimulating (M3) the  
 CC secretion of IL-6 (interleukin-6) in a mammal comprising administering TS  
 CC or an IL-6 inducing variant. The fusion partner comprises a mammalian  
 CC neurotrophic factor which is ciliary neurotrophic factor (CNTF) or  
 CC leukaemia inhibitory factor (LIF). The peptides are useful in providing  
 CC trophic support for neurons and glial cells in a mammal suffering a  
 CC condition selected from: amyotrophic lateral sclerosis, Alzheimer's  
 CC disease, Parkinson's disease, Huntington's disease, Chagas' disease,  
 CC peripheral neuropathy, palsies, multiple sclerosis, stroke, brain trauma,  
 CC spinal cord trauma and peripheral nerve trauma, and in stimulating the  
 CC secretion of IL-6. The present sequence represents a trans-sialidase  
 CC neurotrophic peptide.  
 XX  
 SQ Sequence 45 AA;

Query Match 86.5%; Score 67.5; DB 7; Length 45;  
 Best Local Similarity 93.3%; Pred. No. 0.00021;  
 Matches 14; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 RQRLP-KRMGGSYRC 14  
 |||||  
 Db 24 RQRLPLKRMGGSYRC 38

RESULT 5  
 AAW89049  
 ID AAW89049 standard; protein; 30 AA.  
 XX  
 AC AAW89049;  
 XX  
 DT 01-MAR-1999 (first entry)  
 XX  
 DE Polypeptide fragment encoded by gene 183.  
 XX  
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9854963-A2.  
 XX  
 PD 10-DEC-1998.  
 XX  
 PF 04-JUN-1998; 98WO-US011422.  
 XX  
 PR 06-JUN-1997; 97US-0048875P.  
 PR 06-JUN-1997; 97US-0048876P.  
 PR 06-JUN-1997; 97US-0048877P.  
 PR 06-JUN-1997; 97US-0048878P.  
 PR 06-JUN-1997; 97US-0048880P.  
 PR 06-JUN-1997; 97US-0048881P.  
 PR 06-JUN-1997; 97US-0048882P.  
 PR 06-JUN-1997; 97US-0048883P.  
 PR 06-JUN-1997; 97US-0048884P.  
 PR 06-JUN-1997; 97US-0048885P.  
 PR 06-JUN-1997; 97US-0048892P.  
 PR 06-JUN-1997; 97US-0048893P.  
 PR 06-JUN-1997; 97US-0048894P.  
 PR 06-JUN-1997; 97US-0048895P.  
 PR 06-JUN-1997; 97US-0048896P.  
 PR 06-JUN-1997; 97US-0048897P.  
 PR 06-JUN-1997; 97US-0048898P.  
 PR 06-JUN-1997; 97US-0048899P.  
 PR 06-JUN-1997; 97US-0048900P.  
 PR 06-JUN-1997; 97US-0048901P.



PI Florence C, Hu J, Li Y, Kyaw H, Fischer CL, Ferrie AM, Fan P;  
PI Feng P, Endress GA, Dillon PJ, Carter KC, Brewer LA, Yu G, Zeng Z;  
PI Greene JW;  
XX WPI; 2001-625724/72.  
XX Nucleic acids encoding 207 human secreted polypeptides, useful for  
PT preventing, diagnosing and/or treating, e.g. cancers, Parkinson's disease  
PT and diabetic retinopathy.  
XX Disclosure; Page 398; 1533pp; English.  
XX ABB50301 to ABB51287 and ABA83194 to ABA83441 represent human secreted  
CC proteins (I) and polynucleotide (II) sequences. (I) and (II) have various  
CC activities based on the tissues and cells the genes are expressed in.  
CC Example of these activities include: immunomodulatory; antisclerotic;  
CC dermatological; immunosuppressive; antiinflammatory; immunostimulant;  
CC anti-HIV; cytostatic; cardiac; anti-angiogenic; ophthalmological;  
CC neuroprotective; nootropic; anticonvulsant; antialzheimers; vascular;  
CC antiparkinsonian; antimicrobial; and vulnary. (I) and (II) can be used  
CC in gene therapy and vaccine production. (I) and (II) can be used in the  
CC prevention, diagnosis and treatment of immune disorders (e.g. multiple  
CC sclerosis, systemic lupus erythematosus and human immunodeficiency virus  
CC (HIV) infections), hyperproliferative disorders (e.g. cancers and  
CC Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome,  
CC Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic  
CC disorders (e.g. corneal graft neovascularisation and diabetic  
CC retinopathy), neurological disorders (e.g. Huntington's chorea,  
CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or  
CC for promoting wound healing, regeneration and/or chemotaxis. ABA83185 to  
CC ABA83193 and ABB50300 represent sequences used in the exemplification of  
CC the present invention  
XX Sequence 30 AA;  
SQ Query Match 53.8%; Score 42; DB 4; Length 30;  
Best Local Similarity 61.5%; Pred. No. 4.7;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 RQRLPKRMGGSYR 13  
Db ||| | : | : | : |  
5 RQRRPRRRGGTSR 17  
RESULT 7  
ABO45471  
ID ABO45471 standard; protein; 30 AA.  
XX AC ABO45471;  
XX DT 03-OCT-2003 (first entry)  
XX DE Novel human secreted protein #183 fragment #3.  
XX Human; gene therapy; autoimmune disorder; multiple sclerosis; cancer;  
KW systemic lupus erythematosus; haematopoietic cell disorder; allergy;  
KW agammaglobulinaemia; ataxia telangiectasia; blood coagulation disorder;  
KW afibrinogenemia; thrombocytopenia; graft-versus-host disease; arthritis;  
KW inflammatory condition; ischaemia-reperfusion injury; infectious disease;  
KW hyperproliferative disorder; purpura; viral infection; regeneration;  
KW bacterial infection; ulcer; Alzheimer's disease.  
XX Homo sapiens.  
XX OS  
XX US2003065160-A1.  
XX PN  
XX PD 03-APR-2003.  
XX PP 07-DEC-2001; 2001US-00004860.  
XX 06-JUN-1997; 97US-0048875P.  
PR 06-JUN-1997; 97US-0048876P.  
PR 06-JUN-1997; 97US-0048877P.  
PR 06-JUN-1997; 97US-0048878P.  
PR 06-JUN-1997; 97US-0048879P.  
PR 06-JUN-1997; 97US-0048880P.  
PR 06-JUN-1997; 97US-0048881P.  
PR 06-JUN-1997; 97US-0048882P.  
PR 06-JUN-1997; 97US-0048883P.  
PR 06-JUN-1997; 97US-0048884P.  
PR 06-JUN-1997; 97US-0048885P.  
PR 06-JUN-1997; 97US-0048886P.  
PR 06-JUN-1997; 97US-0048887P.  
PR 06-JUN-1997; 97US-0048888P.  
PR 06-JUN-1997; 97US-0048889P.  
PR 06-JUN-1997; 97US-0048890P.  
PR 06-JUN-1997; 97US-0048891P.  
PR 06-JUN-1997; 97US-0048892P.  
PR 06-JUN-1997; 97US-0048893P.  
PR 06-JUN-1997; 97US-0048894P.  
PR 06-JUN-1997; 97US-0048895P.  
PR 06-JUN-1997; 97US-0048896P.  
PR 06-JUN-1997; 97US-0048897P.  
PR 06-JUN-1997; 97US-0048898P.  
PR 06-JUN-1997; 97US-0048899P.  
PR 06-JUN-1997; 97US-0048900P.  
PR 06-JUN-1997; 97US-0048901P.  
PR 06-JUN-1997; 97US-0048902P.  
PR 06-JUN-1997; 97US-0048903P.  
PR 06-JUN-1997; 97US-0048904P.  
PR 06-JUN-1997; 97US-0048905P.  
PR 06-JUN-1997; 97US-0048906P.  
PR 06-JUN-1997; 97US-0048907P.  
PR 06-JUN-1997; 97US-0048908P.  
PR 06-JUN-1997; 97US-0048909P.  
PR 06-JUN-1997; 97US-0048910P.  
PR 06-JUN-1997; 97US-0048911P.  
PR 06-JUN-1997; 97US-0048912P.  
PR 06-JUN-1997; 97US-0048913P.  
PR 06-JUN-1997; 97US-0048914P.  
PR 06-JUN-1997; 97US-0048915P.  
PR 06-JUN-1997; 97US-0048916P.  
PR 06-JUN-1997; 97US-0048917P.  
PR 06-JUN-1997; 97US-0048918P.  
PR 06-JUN-1997; 97US-0048919P.  
PR 06-JUN-1997; 97US-0048920P.  
PR 06-JUN-1997; 97US-0048921P.  
PR 06-JUN-1997; 97US-0048922P.  
PR 06-JUN-1997; 97US-0048923P.  
PR 06-JUN-1997; 97US-0048924P.  
PR 06-JUN-1997; 97US-0048925P.  
PR 06-JUN-1997; 97US-0048926P.  
PR 06-JUN-1997; 97US-0048927P.  
PR 06-JUN-1997; 97US-0048928P.  
PR 06-JUN-1997; 97US-0048929P.  
PR 06-JUN-1997; 97US-0048930P.  
PR 06-JUN-1997; 97US-0048931P.  
PR 06-JUN-1997; 97US-0048932P.  
PR 06-JUN-1997; 97US-0048933P.  
PR 06-JUN-1997; 97US-0048934P.  
PR 06-JUN-1997; 97US-0048935P.  
PR 06-JUN-1997; 97US-0048936P.  
PR 06-JUN-1997; 97US-0048937P.  
PR 06-JUN-1997; 97US-0048938P.  
PR 06-JUN-1997; 97US-0048939P.  
PR 06-JUN-1997; 97US-0048940P.  
PR 06-JUN-1997; 97US-0048941P.  
PR 06-JUN-1997; 97US-0048942P.  
PR 06-JUN-1997; 97US-0048943P.  
PR 06-JUN-1997; 97US-0048944P.  
PR 06-JUN-1997; 97US-0048945P.  
PR 06-JUN-1997; 97US-0048946P.  
PR 06-JUN-1997; 97US-0048947P.  
PR 06-JUN-1997; 97US-0048948P.  
PR 06-JUN-1997; 97US-0048949P.  
PR 06-JUN-1997; 97US-0048950P.  
PR 06-JUN-1997; 97US-0048951P.  
PR 06-JUN-1997; 97US-0048952P.  
PR 06-JUN-1997; 97US-0048953P.  
PR 06-JUN-1997; 97US-0048954P.  
PR 06-JUN-1997; 97US-0048955P.  
PR 06-JUN-1997; 97US-0048956P.  
PR 06-JUN-1997; 97US-0048957P.  
PR 06-JUN-1997; 97US-0048958P.  
PR 06-JUN-1997; 97US-0048959P.  
PR 06-JUN-1997; 97US-0048960P.  
PR 06-JUN-1997; 97US-0048961P.  
PR 06-JUN-1997; 97US-0048962P.  
PR 06-JUN-1997; 97US-0048963P.  
PR 06-JUN-1997; 97US-0048964P.  
PR 06-JUN-1997; 97US-0048965P.  
PR 06-JUN-1997; 97US-0048966P.  
PR 06-JUN-1997; 97US-0048967P.  
PR 06-JUN-1997; 97US-0048968P.  
PR 06-JUN-1997; 97US-0048969P.  
PR 06-JUN-1997; 97US-0048970P.  
PR 06-JUN-1997; 97US-0048971P.  
PR 06-JUN-1997; 97US-0048972P.  
PR 06-JUN-1997; 97US-0048973P.  
PR 06-JUN-1997; 97US-0048974P.  
PR 06-JUN-1997; 97US-0048975P.  
PR 06-JUN-1997; 97US-0048976P.  
PR 06-JUN-1997; 97US-0048977P.  
PR 06-JUN-1997; 97US-0048978P.  
PR 06-JUN-1997; 97US-0048979P.  
PR 06-JUN-1997; 97US-0048980P.  
PR 06-JUN-1997; 97US-0048981P.  
PR 06-JUN-1997; 97US-0048982P.  
PR 06-JUN-1997; 97US-0048983P.  
PR 06-JUN-1997; 97US-0048984P.  
PR 06-JUN-1997; 97US-0048985P.  
PR 06-JUN-1997; 97US-0048986P.  
PR 06-JUN-1997; 97US-0048987P.  
PR 06-JUN-1997; 97US-0048988P.  
PR 06-JUN-1997; 97US-0048989P.  
PR 06-JUN-1997; 97US-0048990P.  
PR 06-JUN-1997; 97US-0048991P.  
PR 06-JUN-1997; 97US-0048992P.  
PR 06-JUN-1997; 97US-0048993P.  
PR 06-JUN-1997; 97US-0048994P.  
PR 06-JUN-1997; 97US-0048995P.  
PR 06-JUN-1997; 97US-0048996P.  
PR 06-JUN-1997; 97US-0048997P.  
PR 06-JUN-1997; 97US-0048998P.  
PR 06-JUN-1997; 97US-0048999P.  
PR 06-JUN-1997; 97US-0049000P.  
PR 06-JUN-1997; 97US-0049001P.  
PR 06-JUN-1997; 97US-0049002P.  
PR 06-JUN-1997; 97US-0049003P.  
PR 06-JUN-1997; 97US-0049004P.  
PR 06-JUN-1997; 97US-0049005P.  
PR 06-JUN-1997; 97US-0049006P.  
PR 06-JUN-1997; 97US-0049007P.  
PR 06-JUN-1997; 97US-0049008P.  
PR 06-JUN-1997; 97US-0049009P.  
PR 06-JUN-1997; 97US-0049010P.  
PR 06-JUN-1997; 97US-0049011P.  
PR 06-JUN-1997; 97US-0049012P.  
PR 06-JUN-1997; 97US-0049013P.  
PR 06-JUN-1997; 97US-0049014P.  
PR 06-JUN-1997; 97US-0049015P.  
PR 06-JUN-1997; 97US-0049016P.  
PR 06-JUN-1997; 97US-0049017P.  
PR 06-JUN-1997; 97US-0049018P.  
PR 06-JUN-1997; 97US-0049019P.  
PR 06-JUN-1997; 97US-0049020P.  
PR 06-JUN-1997; 97US-0049021P.  
PR 06-JUN-1997; 97US-0049022P.  
PR 06-JUN-1997; 97US-0049023P.  
PR 06-JUN-1997; 97US-0049024P.  
PR 06-JUN-1997; 97US-0049025P.  
PR 06-JUN-1997; 97US-0049026P.  
PR 06-JUN-1997; 97US-0049027P.  
PR 06-JUN-1997; 97US-0049028P.  
PR 06-JUN-1997; 97US-0049029P.  
PR 06-JUN-1997; 97US-0049030P.  
PR 06-JUN-1997; 97US-0049031P.  
PR 06-JUN-1997; 97US-0049032P.  
PR 06-JUN-1997; 97US-0049033P.  
PR 06-JUN-1997; 97US-0049034P.  
PR 06-JUN-1997; 97US-0049035P.  
PR 06-JUN-1997; 97US-0049036P.  
PR 06-JUN-1997; 97US-0049037P.  
PR 06-JUN-1997; 97US-0049038P.  
PR 06-JUN-1997; 97US-0049039P.  
PR 06-JUN-1997; 97US-0049040P.  
PR 06-JUN-1997; 97US-0049041P.  
PR 06-JUN-1997; 97US-0049042P.  
PR 06-JUN-1997; 97US-0049043P.  
PR 06-JUN-1997; 97US-0049044P.  
PR 06-JUN-1997; 97US-0049045P.  
PR 06-JUN-1997; 97US-0049046P.  
PR 06-JUN-1997; 97US-0049047P.  
PR 06-JUN-1997; 97US-0049048P.  
PR 06-JUN-1997; 97US-0049049P.  
PR 06-JUN-1997; 97US-0049050P.  
PR 06-JUN-1997; 97US-0049051P.  
PR 06-JUN-1997; 97US-0049052P.  
PR 06-JUN-1997; 97US-0049053P.  
PR 06-JUN-1997; 97US-0049054P.  
PR 06-JUN-1997; 97US-0049055P.  
PR 06-JUN-1997; 97US-0049056P.  
PR 06-JUN-1997; 97US-0049057P.  
PR 06-JUN-1997; 97US-0049058P.  
PR 06-JUN-1997; 97US-0049059P.  
PR 06-JUN-1997; 97US-0049060P.  
PR 06-JUN-1997; 97US-0049061P.  
PR 06-JUN-1997; 97US-0049062P.  
PR 06-JUN-1997; 97US-0049063P.  
PR 06-JUN-1997; 97US-0049064P.  
PR 06-JUN-1997; 97US-0049065P.  
PR 06-JUN-1997; 97US-0049066P.  
PR 06-JUN-1997; 97US-0049067P.  
PR 06-JUN-1997; 97US-0049068P.  
PR 06-JUN-1997; 97US-0049069P.  
PR 06-JUN-1997; 97US-0049070P.  
PR 06-JUN-1997; 97US-0049071P.  
PR 06-JUN-1997; 97US-0049072P.  
PR 06-JUN-1997; 97US-0049073P.  
PR 06-JUN-1997; 97US-0049074P.  
PR 06-JUN-1997; 97US-0049075P.  
PR 06-JUN-1997; 97US-0049076P.  
PR 06-JUN-1997; 97US-0049077P.  
PR 06-JUN-1997; 97US-0049078P.  
PR 06-JUN-1997; 97US-0049079P.  
PR 06-JUN-1997; 97US-0049080P.  
PR 06-JUN-1997; 97US-0049081P.  
PR 06-JUN-1997; 97US-0049082P.  
PR 06-JUN-1997; 97US-0049083P.  
PR 06-JUN-1997; 97US-0049084P.  
PR 06-JUN-1997; 97US-0049085P.  
PR 06-JUN-1997; 97US-0049086P.  
PR 06-JUN-1997; 97US-0049087P.  
PR 06-JUN-1997; 97US-0049088P.  
PR 06-JUN-1997; 97US-0049089P.  
PR 06-JUN-1997; 97US-0049090P.  
PR 06-JUN-1997; 97US-0049091P.  
PR 06-JUN-1997; 97US-0049092P.  
PR 06-JUN-1997; 97US-0049093P.  
PR 06-JUN-1997; 97US-0049094P.  
PR 06-JUN-1997; 97US-0049095P.  
PR 06-JUN-1997; 97US-0049096P.  
PR 06-JUN-1997; 97US-0049097P.  
PR 06-JUN-1997; 97US-0049098P.  
PR 06-JUN-1997; 97US-0049099P.  
PR 06-JUN-1997; 97US-0049100P.  
PR 06-JUN-1997; 97US-0049101P.  
PR 06-JUN-1997; 97US-0049102P.  
PR 06-JUN-1997; 97US-0049103P.  
PR 06-JUN-1997; 97US-0049104P.  
PR 06-JUN-1997; 97US-0049105P.  
PR 06-JUN-1997; 97US-0049106P.  
PR 06-JUN-1997; 97US-0049107P.  
PR 06-JUN-1997; 97US-0049108P.  
PR 06-JUN-1997; 97US-0049109P.  
PR 06-JUN-1997; 97US-0049110P.  
PR 06-JUN-1997; 97US-0049111P.  
PR 06-JUN-1997; 97US-0049112P.  
PR 06-JUN-1997; 97US-0049113P.  
PR 06-JUN-1997; 97US-0049114P.  
PR 06-JUN-1997; 97US-0049115P.  
PR 06-JUN-1997; 97US-0049116P.  
PR 06-JUN-1997; 97US-0049117P.  
PR 06-JUN-1997; 97US-0049118P.  
PR 06-JUN-1997; 97US-0049119P.  
PR 06-JUN-1997; 97US-0049120P.  
PR 06-JUN-1997; 97US-0049121P.  
PR 06-JUN-1997; 97US-0049122P.  
PR 06-JUN-1997; 97US-0049123P.  
PR 06-JUN-1997; 97US-0049124P.  
PR 06-JUN-1997; 97US-0049125P.  
PR 06-JUN-1997; 97US-0049126P.  
PR 06-JUN-1997; 97US-0049127P.  
PR 06-JUN-1997; 97US-0049128P.  
PR 06-JUN-1997; 97US-0049129P.  
PR 06-JUN-1997; 97US-0049130P.  
PR 06-JUN-1997; 97US-0049131P.  
PR 06-JUN-1997; 97US-0049132P.  
PR 06-JUN-1997; 97US-0049133P.  
PR 06-JUN-1997; 97US-0049134P.  
PR 06-JUN-1997; 97US-0049135P.  
PR 06-JUN-1997; 97US-0049136P.  
PR 06-JUN-1997; 97US-0049137P.  
PR 06-JUN-1997; 97US-0049138P.  
PR 06-JUN-1997; 97US-0049139P.  
PR 06-JUN-1997; 97US-0049140P.  
PR 06-JUN-1997; 97US-0049141P.  
PR 06-JUN-1997; 97US-0049142P.  
PR 06-JUN-1997; 97US-0049143P.  
PR 06-JUN-1997; 97US-0049144P.  
PR 06-JUN-1997; 97US-0049145P.  
PR 06-JUN-1997; 97US-0049146P.  
PR 06-JUN-1997; 97US-0049147P.  
PR 06-JUN-1997; 97US-0049148P.  
PR 06-JUN-1997; 97US-0049149P.  
PR 06-JUN-1997; 97US-0049150P.  
PR 06-JUN-1997; 97US-0049151P.  
PR 06-JUN-1997; 97US-0049152P.  
PR 06-JUN-1997; 97US-0049153P.  
PR 06-JUN-1997; 97US-0049154P.  
PR 06-JUN-1997; 97US-0049155P.  
PR 06-JUN-1997; 97US-0049156P.  
PR 06-JUN-1997; 97US-0049157P.  
PR 06-JUN-1997; 97US-0049158P.  
PR 06-JUN-1997; 97US-0049159P.  
PR 06-JUN-1997; 97US-0049160P.  
PR 06-JUN-1997; 97US-0049161P.  
PR 06-JUN-1997; 97US-0049162P.  
PR 06-JUN-1997; 97US-0049163P.  
PR 06-JUN-1997; 97US-0049164P.  
PR 06-JUN-1997; 97US-0049165P.  
PR 06-JUN-1997; 97US-0049166P.  
PR 06-JUN-1997; 97US-0049167P.  
PR 06-JUN-1997; 97US-0049168P.  
PR 06-JUN-1997; 97US-0049169P.  
PR 06-JUN-1997; 97US-0049170P.  
PR 06-JUN-1997; 97US-0049171P.  
PR 06-JUN-1997; 97US-0049172P.  
PR 06-JUN-1997; 97US-0049173P.  
PR 06-JUN-1997; 97US-0049174P.  
PR 06-JUN-1997; 97US-0049175P.  
PR 06-JUN-1997; 97US-0049176P.  
PR 06-JUN-1997; 97US-0049177P.  
PR 06-JUN-1997; 97US-0049178P.  
PR 06-JUN-1997; 97US-0049179P.  
PR 06-JUN-1997; 97US-0049180P.  
PR 06-JUN-1997; 97US-0049181P.  
PR 06-JUN-1997; 97US-0049182P.  
PR 06-JUN-1997; 97US-0049183P.  
PR 06-JUN-1997; 97US-0049184P.  
PR 06-JUN-1997; 97US-0049185P.  
PR 06-JUN-1997; 97US-0049186P.  
PR 06-JUN-1997; 97US-0049187P.  
PR 06-JUN-1997; 97US-0049188P.  
PR 06-JUN-1997; 97US-0049189P.  
PR 06-JUN-1997; 97US-0049190P.  
PR 06-JUN-1997; 97US-0049191P.  
PR 06-JUN-1997; 97US-0049192P.  
PR 06-JUN-1997; 97US-0049193P.  
PR 06-JUN-1997; 97US-0049194P.  
PR 06-JUN-1997; 97US-0049195P.  
PR 06-JUN-1997; 97US-0049196P.  
PR 06-JUN-1997; 97US-0049197P.  
PR 06-JUN-1997; 97US-0049198P.  
PR 06-JUN-1997; 97US-0049199P.  
PR 06-JUN-1997; 97US-0049200P.  
PR 06-JUN-1997; 97US-0049201P.  
PR 06-JUN-1997; 97US-0049202P.  
PR 06-JUN-1997; 97US-0049203P.  
PR 06-JUN-1997; 97US-0049204P.  
PR 06-JUN-1997; 97US-0049205P.  
PR 06-JUN-1997; 97US-0049206P.  
PR 06-JUN-1997; 97US-0049207P.  
PR 06-JUN-1997; 97US-0049208P.  
PR 06-JUN-1997; 97US-0049209P.  
PR 06-JUN-1997; 97US-0049210P.  
PR 06-JUN-1997; 97US-0049211P.  
PR 06-JUN-1997; 97US-0049212P.  
PR 06-JUN-1997; 97US-0049213P.  
PR 06-JUN-1997; 97US-0049214P.  
PR 06-JUN-1997; 97US-0049215P.  
PR 06-JUN-1997; 97US-0049216P.  
PR 06-JUN-1997; 97US-0049217P.  
PR 06-JUN-1997; 97US-0049218P.  
PR 06-JUN-1997; 97US-0049219P.  
PR 06-JUN-1997; 97US-0049220P.  
PR 06-JUN-1997; 97US-0049221P.  
PR 06-JUN-1997; 97US-0049222P.  
PR 06-JUN-1997; 97US-0049223P.  
PR 06-JUN-1997; 97US-0049224P.  
PR 06-JUN-1997; 97US-0049225P.  
PR 06-JUN-1997; 97US-0049226P.  
PR 06-JUN-1997; 97US-0049227P.  
PR 06-JUN-1997; 97US-0049228P.  
PR 06-JUN-1997; 97US-0049229P.  
PR 06-JUN-1997; 97US-0049230P.  
PR 06-JUN-1997; 97US-0049231P.  
PR 06-JUN-1997; 97US-0049232P.  
PR 06-JUN-1997; 97US-0049233P.  
PR 06-JUN-1997; 97US-0049234P.  
PR 06-JUN-1997; 97US-0049235P.  
PR 06-JUN-1997; 97US-0049236P.  
PR 06-JUN-1997; 97US-0049237P.  
PR 06-JUN-1997; 97US-0049238P.  
PR 06-JUN-1997; 97US-0049239P.  
PR 06-JUN-1997; 97US-0049240P.  
PR 06-JUN-1997; 97US-0049241P.  
PR 06-JUN-1997; 97US-0049242P.  
PR 06-JUN-1997; 97US-0049243P.  
PR 06-JUN-1997; 97US-0049244P.  
PR 06-JUN-1997; 97US-0049245P.  
PR 06-JUN-1997; 97US-0049246P.  
PR 06-JUN-1997; 97US-0049247P.  
PR 06-JUN-1997; 97US-0049248P.  
PR 06-JUN-1997; 97US-0049249P.  
PR 06-JUN-1997; 97US-0049250P.  
PR 06-JUN-1997; 97US-0049251P.  
PR 06-JUN-1997; 97US-0049252P.  
PR 06-JUN-1997; 97US-0049253P.  
PR 06-JUN-1997; 97US-0049254P.  
PR 06-JUN-1997; 97US-0049255P.  
PR 06-JUN-1997; 97US-0049256P.  
PR 06-JUN-1997; 97US-0049257P.  
PR 06-JUN-1997; 97US-0049258P.  
PR 06-JUN-1997; 97US-0049259P.  
PR 06-JUN-1997; 97US-0049260P.  
PR 06

PR 04-DEC-1998; 98US-00205258.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Young P, Greene JM, Ferrie AM, Ruben SM, Rosen CA, Hu J;  
PI Olsen HS, Ebner R, Brewer LA, Moore PA, Shi Y, Florence C;  
PI Florence K, Lafleur DW, Ni J, Fan P, Wei Y, Fischer CL, Soppet DR;  
PI Li Y, Zeng Z, Kyaw H, Yu G, Feng P, Dillion PJ, Endress GA,  
PI Carter KC;  
XX  
XX WPI; 2003-540804/51.  
DR  
XX  
XX New isolated protein, useful for preparing a composition for diagnosing  
PT or treating cancer, inflammatory, immune or infectious diseases.  
XX  
XX Disclosure; Page 112; 172pp; English.  
XX  
XX The invention relates to an isolated HEMAE80 protein. The protein is  
CC useful for preparing a composition for diagnosing or treating autoimmune  
CC disorders e.g. multiple sclerosis and systemic lupus erythematosus;  
CC haematopoietic cell disorders e.g. agammaglobulinaemia and ataxia  
CC telangiectasia; blood coagulation disorders e.g. afibrinogenemia and  
CC thrombocytopenia; allergy; graft-versus-host disease; inflammatory  
CC conditions e.g. ischaemia-reperfusion injury and arthritis;  
CC hyperproliferative disorders e.g. cancer and purpura; infectious disease  
CC e.g. viral infection and bacterial infection. The polynucleotide or  
CC protein can be used to regenerate damaged tissue e.g. ulcers and  
CC Alzheimer's disease. The present sequence represents the amino acid  
CC sequence of a novel human secreted protein fragment. Note: the sequence  
CC data for this patent did not form part of the printed specification but  
CC was obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?docID=20030065160  
XX  
XX Sequence 30 AA;  
SQ

Query March 53.8%; Score 42; DB 6; Length 30;  
Best Local Similarity 61.5%; Pred. No. 4.7;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
OY 1 RQRLPKRMGSGYSR 13  
||| | | | | |  
DB 5 RQRRPRRGTSR 17  
||| | | | | |

RESULT 8  
ABO26951  
ID ABO26951 standard; protein; 30 AA.  
XX  
XX ABO26951;  
AC ABO26951;  
XX  
XX 10-SEP-2003 (first entry)  
XX  
XX Protein associated with novel secreted protein gene 183 #3.  
XX  
XX Secreted protein; precerebellin-like protein; sepsis; acne; psoriasis;  
KW neurodegenerative disorder; behavioural disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; schizophrenia; mania;  
KW dementia; paranoia; psychosis; autism; immune disorder; infection;  
KW inflammation; allergy; liver disorder; hepatoblastoma; jaundice;  
KW hepatitis; immunological disorder; AIDS; leukaemia; rheumatoid arthritis;  
KW cancer.  
XX  
XX Unidentified.  
OS  
XX  
XX US6525174-B1.  
PN  
XX  
XX 25-FEB-2003.  
PD  
XX  
XX 04-DEC-1998; 98US-00205258.  
XX  
XX 06-JUN-1997; 97US-0048875P.  
PR  
XX 06-JUN-1997; 97US-0048876P.  
PR  
XX 06-JUN-1997; 97US-0048877P.  
PR  
XX 06-JUN-1997; 97US-0048878P.  
PR  
XX 06-JUN-1997; 97US-0048879P.  
PR  
XX 06-JUN-1997; 97US-0048880P.  
PR  
XX 06-JUN-1997; 97US-0048881P.  
PR  
XX 06-JUN-1997; 97US-0048882P.  
PR  
XX 06-JUN-1997; 97US-0048883P.  
PR  
XX 06-JUN-1997; 97US-0048884P.  
PR  
XX 06-JUN-1997; 97US-0048885P.  
PR  
XX 06-JUN-1997; 97US-0048892P.  
PR  
XX 06-JUN-1997; 97US-0048893P.  
PR  
XX 06-JUN-1997; 97US-0048894P.  
PR  
XX 06-JUN-1997; 97US-0048895P.  
PR  
XX 06-JUN-1997; 97US-0048896P.  
PR  
XX 06-JUN-1997; 97US-0048897P.  
PR  
XX 06-JUN-1997; 97US-0048898P.  
PR  
XX 06-JUN-1997; 97US-0048899P.  
PR  
XX 06-JUN-1997; 97US-0048900P.  
PR  
XX 06-JUN-1997; 97US-0048901P.  
PR  
XX 06-JUN-1997; 97US-0048915P.  
PR  
XX 06-JUN-1997; 97US-0048916P.  
PR  
XX 06-JUN-1997; 97US-0048917P.  
PR  
XX 06-JUN-1997; 97US-0048949P.  
PR  
XX 06-JUN-1997; 97US-0048962P.  
PR  
XX 06-JUN-1997; 97US-0048963P.  
PR  
XX 06-JUN-1997; 97US-0048964P.  
PR  
XX 06-JUN-1997; 97US-0048970P.  
PR  
XX 06-JUN-1997; 97US-0048971P.  
PR  
XX 06-JUN-1997; 97US-0048972P.  
PR  
XX 06-JUN-1997; 97US-0048974P.  
PR  
XX 06-JUN-1997; 97US-0049019P.  
PR  
XX 06-JUN-1997; 97US-0049020P.  
PR  
XX 06-JUN-1997; 97US-0049373P.  
PR  
XX 06-JUN-1997; 97US-0049374P.  
PR  
XX 06-JUN-1997; 97US-0049375P.  
PR  
XX 05-SEP-1997; 97US-0057584P.  
PR  
XX 05-SEP-1997; 97US-0057627P.  
PR  
XX 05-SEP-1997; 97US-0057628P.  
PR  
XX 05-SEP-1997; 97US-0057629P.  
PR  
XX 05-SEP-1997; 97US-0057634P.  
PR  
XX 05-SEP-1997; 97US-0057635P.  
PR  
XX 05-SEP-1997; 97US-0057642P.  
PR  
XX 05-SEP-1997; 97US-0057643P.  
PR  
XX 05-SEP-1997; 97US-0057644P.  
PR  
XX 05-SEP-1997; 97US-0057645P.  
PR  
XX 05-SEP-1997; 97US-0057646P.  
PR  
XX 05-SEP-1997; 97US-0057647P.  
PR  
XX 05-SEP-1997; 97US-0057648P.  
PR  
XX 05-SEP-1997; 97US-0057649P.  
PR  
XX 05-SEP-1997; 97US-0057650P.  
PR  
XX 05-SEP-1997; 97US-0057651P.  
PR  
XX 05-SEP-1997; 97US-0057654P.  
PR  
XX 05-SEP-1997; 97US-0057661P.  
PR  
XX 05-SEP-1997; 97US-0057662P.  
PR  
XX 05-SEP-1997; 97US-0057666P.  
PR  
XX 05-SEP-1997; 97US-0057667P.  
PR  
XX 05-SEP-1997; 97US-0057668P.  
PR  
XX 05-SEP-1997; 97US-0057760P.  
PR  
XX 05-SEP-1997; 97US-0057761P.  
PR  
XX 05-SEP-1997; 97US-0057762P.  
PR  
XX 05-SEP-1997; 97US-0057763P.  
PR  
XX 05-SEP-1997; 97US-0057764P.  
PR  
XX 05-SEP-1997; 97US-0057765P.  
PR  
XX 05-SEP-1997; 97US-0057769P.  
PR  
XX 05-SEP-1997; 97US-0057770P.  
PR  
XX 05-SEP-1997; 97US-0057771P.  
PR  
XX 05-SEP-1997; 97US-0057774P.  
PR  
XX 05-SEP-1997; 97US-0057775P.  
PR  
XX 05-SEP-1997; 97US-0057776P.  
PR  
XX 05-SEP-1997; 97US-0057777P.  
PR  
XX 05-SEP-1997; 97US-0057778P.  
PR  
XX 18-DEC-1997; 97US-0070923P.  
PR  
XX 04-JUN-1998; 98WO-US011422.  
PR  
XX 15-JUL-1998; 98US-0092921P.  
PR  
XX 30-JUL-1998; 98US-0094657P.  
PR

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX Young P, Greene JM, Ferrie AM, Ruben SM, Rosen CA, Hu J;  
 PI Olsen HS, Ebner R, Brewer LA, Moore PA, Shi Y, Florence C;  
 PI Florence K, Lafleur DM, Ni J, Fan P, Wei Y, Fischer CL, Soppet DR;  
 PI Li Y, Zeng Z, Kyaw H, Yu G, Feng P, Dillon PJ, Endress GA;  
 PI Carter KC;

XX WPI; 2003-511926/48.

XX New precerebellin-like protein, useful for diagnosing or treating  
 PT neurodegenerative and behavioral disorders, immune disorders, liver  
 PT disorders, and cancer.

XX Disclosure; Col 203-204; 156pp; English.

XX The invention relates to an isolated protein comprising amino acid  
 CC residues 33-205 or 1-205 of a novel human secreted protein appearing as  
 CC ABO26252. The protein is encoded by one of 238 disclosed cDNA sequences  
 CC encoding 238 secreted proteins. ABO26252 is a precerebellin-like protein.  
 CC Also included are a composition comprising the protein and a carrier and  
 CC an isolated protein produced by expressing the protein cited above by a  
 CC cell, and recovering the protein. The proteins are useful for diagnosing  
 CC or treating neurodegenerative and behavioural disorders (e.g. Alzheimer's  
 CC disease, Parkinson's disease, Huntington's disease, schizophrenia, mania,  
 CC dementia, paranoia, psychoses or autism), immune disorders (e.g.  
 CC infection, inflammation, allergy), liver disorders (e.g. hepatoblastoma,  
 CC jaundice, hepatitis), immunological disorders (e.g. AIDS, leukaemia,  
 CC rheumatoid arthritis, sepsis, acne, psoriasis) and cancer. The present  
 CC sequence is a protein associated with one of the 238 disclosed novel  
 CC secreted proteins

XX Sequence 30 AA;

Query Match 53.8%; Score 42; DB 7; Length 30;  
 Best Local Similarity 61.5%; Pred. No. 4.7;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGGSYR 13  
 ||| ||| |||  
 Db 5 RQRRPRRGGTGR 17

RESULT 9  
 ABB68550  
 ID ABB68550 standard; protein; 339 AA.

XX AC ABB68550;

XX DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 32442.

XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL12653.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

XX Disclosure; SEQ ID NO 32442; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 339 AA;

Query Match 53.8%; Score 42; DB 4; Length 339;  
 Best Local Similarity 70.0%; Pred. No. 57;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 5 PKRMGGSYR 14  
 ||| ||| |||  
 Db 208 PKRVPGGYR 217

RESULT 10  
 AAU42341  
 ID AAU42341 standard; protein; 189 AA.

XX AC AAU42341;

XX DT 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #3237.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.

XX 02-JUN-2000; 2000US-0208841P.

XX 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS59516.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris.

XX Example 1; SEQ ID NO 3536; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in

the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA). Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 189 AA;

Query Match 52.6%; Score 41; DB 4; Length 189;  
 Best Local Similarity 61.5%; Pred. No. 47;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGGYSR 13  
 ||||| : ||:  
 Db 177 RQRLPPQAGSHR 189

RESULT 11  
 ABM38860  
 ID ABM38860 standard; protein; 189 AA.  
 AC ABM38860;  
 DT 20-OCT-2003 (first entry)  
 DE Propionibacterium acnes predicted ORF-encoded polypeptide #3536.  
 KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
 KW immunostimulant; immune response; vaccine.  
 OS Propionibacterium acnes.  
 XX WO2003033515-A1.  
 XX 24-APR-2003.  
 XX 11-OCT-2002; 2002WO-US032727.  
 XX 15-OCT-2001; 2001US-00978025.  
 XX (CORI-) CORIXA CORP.  
 XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
 FI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 PI Barth B, Vallieve-Douglass J;  
 XX WPI; 2003-381789/36.  
 DR N-PSDB; ACF64445.  
 XX New Propionibacterium acnes polypeptides and polynucleotides encoding the polypeptide, useful for diagnosing, preventing or treating acne vulgaris, or for stimulating an immune response specific for a P. acnes protein.  
 PS Example 1; SEQ ID NO 3536; 1481pp; English.  
 XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733) encoding a Propionibacterium acnes protein. The invention also relates to polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to immunogenic fragments of P. acnes polypeptides. The invention additionally encompasses expression vectors and host cells comprising a

polynucleotide of the invention; antibodies against polypeptides of the invention; fusion proteins comprising a polypeptide of the invention; a method for stimulating an immune response specific for a P. acnes polypeptide and an isolated T cell population comprising T cells prepared via this method; a vaccine composition (comprising P. acnes polypeptides, polynucleotides, antibodies, fusion proteins, T cell populations, or antigen-presenting cells that express the polypeptide); a method and kit for detecting or determining the presence or absence of P. acnes in a patient; and a method for inhibiting the development of P. acnes in a patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion proteins, T cell populations or antigen-presenting cells that express the polypeptides are useful for diagnosing, preventing or treating acne vulgaris, or for stimulating an immune response specific for a P. acnes protein. The polynucleotides can also be used as probes or primers for nucleic acid hybridisation. The vaccine composition is useful for the stimulation of an immune response against P. acnes, or for treating acne, and the kit is useful for performing a diagnostic assay. The present sequence represents a polypeptide predicted to be encoded by an ORF (open reading frame) contained within the P. acnes polynucleotides of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 189 AA;

Query Match 52.6%; Score 41; DB 6; Length 189;  
 Best Local Similarity 61.5%; Pred. No. 47;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGGYSR 13  
 ||||| : ||:  
 Db 177 RQRLPPQAGSHR 189

RESULT 12  
 ABB75629  
 ID ABB75629 standard; protein; 212 AA.  
 AC ABB75629;  
 XX 10-JUN-2002 (first entry)  
 XX Human endo type protease 23.32.  
 XX Endo type protease 23.32; endoprotease; human; tumour; haemopathy;  
 KW HIV infection; immunological disease; inflammation; cytostatic;  
 KW haemostatic; anti-HIV; virucide; immunomodulator; antiinflammatory;  
 KW enzyme; gene therapy.  
 OS Homo sapiens.  
 XX WO200220744-A1.  
 XX 14-MAR-2002.  
 XX 02-JUL-2001; 2001WO-CN001144.  
 XX 07-JUL-2000; 2000CN-00119412.  
 XX (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.  
 XX Mao Y, Xie Y;  
 XX WPI; 2002-269623/31.  
 DR N-PSDB; ABL53561.  
 XX Human endo type protease 23.32 and encoding polynucleotide, used in diagnosis and treatment of malignant tumors, hemopathy, human immunodeficiency virus infection, immunological diseases and inflammation.  
 PS Claim 1; Page 30; 36pp; Chinese.



CC The present sequence is the protein sequence for human endo type protease  
CC 23.32. The protease protein and its coding sequence are useful for the  
CC diagnosis and treatment of malignant tumours, haemopathy, HIV infection,  
CC immunological disease and inflammation

XX SQ Sequence 212 AA;

Query Match 52.6%; Score 41; DB 5; Length 212;  
Best Local Similarity 63.6%; Pred. No. 53;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 RLPKRMGGSYR 13  
|||:|:|  
Db 198 RLPKRMGGNFR 208

RESULT 13  
AAR05328  
ID AAR05328 standard; protein; 235 AA.

XX AC AAR05328;

DT 25-MAR-2003 (revised)  
DT 03-DEC-1990 (first entry)

XX Rat tumour necrosis factor (TNF) gene product.

XX Tumour necrosis factor; TNF; antioncotic; cancer; ds.

OS Rattus sp.

XX JP02157295-A.

XX 18-JUN-1990.

PF 07-DEC-1988; 88JJP-00307751.

PR 07-DEC-1988; 88JJP-00307751.

PA (ASAH ) ASahi CHEM IND CO LTD.

XX WPI; 1990-228715/30.

DR N-PSDB; AAR05328.

XX Novel anti-oncotic polypeptide - prepd. by culturing microbe contg. DNA.

PS Disclosure; Page ?; -pp; Japanese.

CC Fragment of gene product (AAs 80-235) is claimed as an antioncotic  
CC peptide, produced from a transformed microorganism expression system.  
CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 235 AA;

Query Match 51.9%; Score 40.5; DB 2; Length 235;  
Best Local Similarity 56.2%; Pred. No. 72;  
Matches 9; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 2 QRLPKRMGG---SYRC 14  
:|:|:|:|  
Db 15 EALPKRMGGGLQNSRRRC 30

RESULT 14  
AAR05996  
ID AAR05996 standard; protein; 235 AA.

XX AC AAR05996;

DT 25-MAR-2003 (revised)

DT 03-DEC-1990 (first entry)

XX Rat tumour necrosis factor (TNF).

XX Tumour necrosis factor; TNF; antioncotic; cancer; ds.

OS Rattus sp.

XX JP02157295-A.

XX 18-JUN-1990.

PF 07-DEC-1988; 88JJP-00307751.

PR 07-DEC-1988; 88JJP-00307751.

PA (ASAH ) ASahi CHEM IND CO LTD.

XX WPI; 1990-228715/30.

DR P-PSDB; AAR05996.

XX Novel anti-oncotic polypeptide - prepd. by culturing microbe contg. DNA.

PS Disclosure; Page ?; -pp; Japanese.

CC Fragment of gene product (AAs 80-235) is claimed as an antioncotic  
CC peptide, produced from a transformed microorganism expression system.  
CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 235 AA;

Query Match 51.9%; Score 40.5; DB 2; Length 235;  
Best Local Similarity 56.2%; Pred. No. 72;  
Matches 9; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 2 QRLPKRMGG---SYRC 14  
:|:|:|:|  
Db 15 EALPKRMGGGLQNSRRRC 30

RESULT 15

AAU10546

ID AAU10546 standard; protein; 235 AA.

XX AC AAU10546;

XX 14-FEB-2002 (first entry)

DE Rat tumour necrosis factor (TNF) alpha (YSG10) polypeptide.

XX YSG; YSG10; schizophrenia; chronic animal model; LCGU; netrin receptor;  
XX local cerebral glucose utilisation; phosphodiesterase 1-alpha; UNC5H1;  
XX calcium-independent alpha-latrotoxin receptor; CIRL; tRKE; synapsin IA;  
XX epithelial discoidin domain receptor 1; synapsin IB; neuroleptic; ss;  
XX tumour necrosis factor alpha; TNF-alpha; rat.

OS Rattus sp.

Key Location/Qualifiers  
FT Misc-difference 190 /note= "Encoded by GAG"

XX WO200175440-A2.

XX 11-OCT-2001.

XX 02-APR-2001; 2001WO-GB001486.

XX 31-MAR-2000; 2000GB-00007880.

XX 26-MAY-2000; 2000GB-00012768.

XX (WELF-) WELFIDE CORP.

XX Cochran S, Paterson G, Ohashi Y, Morris B, Pratt J;

XX WPI; 2002-010813/01.

DR N-PSDB; AAS16846.  
XX Novel chronic animal model of schizophrenia, useful for identifying anti-  
PT psychotic drugs and genes that are associated with schizophrenia.  
XX  
XX  
XX PS Disclosure; Fig 10b; 79pp; English.  
XX  
XX The invention relates to YSG polynucleotide fragments for use in  
CC diagnosing and/or developing treatments for schizophrenia using chronic  
CC animal models. The polynucleotides and their encoded polypeptides are  
CC used for identification of compounds which modulate the expression of YSG  
CC molecules, leading to the manufacture of schizophrenia medicaments. The  
CC sequences can also be used for testing candidate compounds for any effect  
CC on the polypeptides. Anti-schizophrenic effects of a compound can be  
CC determined by measuring local cerebral glucose utilisation (LCGU) or  
CC comparing its expression level with that of a control group. The  
CC sequences are useful in the identification of genes associated with  
CC schizophrenic states and in the development of an antibody. The sequences  
CC of the invention include phosphodiesterase 1-alpha, calcium-independent  
CC alpha-latrotoxin receptors (CIRL)-1,2&3, epithelial discoidin domain  
CC receptor 1 (trke), netrin receptor (UNC5H1), synapsins 1A and AB and  
CC tumour necrosis factor (TNF) alpha. This sequence represents rat tumour  
CC necrosis factor (TNF) alpha (YSG10) polypeptide  
XX  
XX SQ Sequence 235 AA;

Query Match 51.9%; Score 40.5; DB 5; Length 235;  
Best Local Similarity 56.2%; Pred. No. 72;  
Matches 9; Conservative 2; Mismatches 2; Indels 3; Gaps 1;  
  
QY 2 QRLPKRMGG---SYRC 14  
: |||: |||  
Db 15 EALPKKMGGLQNSRRC 30

Search completed: June 3, 2004, 14:25:50  
Job time : 26.025 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 14:09:38 ; Search time 7.525 Seconds  
(without alignments) 96.048 Million cell updates

Title: US-09-745-008-14  
Perfect score: 78  
Sequence: 1 RQRLPKRMGGSYRC 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp:*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp:*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp:*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp:*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pcp:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Sequence	28695, A
1	45	57.7	433	4	US-09-252-991A-28695	Sequence 28695, A	Sequence 28695, A	
2	43	55.1	164	4	US-09-252-991A-20615	Sequence 20615, A	Sequence 20615, A	
3	43	55.1	164	4	US-09-252-991A-30154	Sequence 30154, A	Sequence 30154, A	
4	43	55.1	223	4	US-09-252-991A-18819	Sequence 18819, A	Sequence 18819, A	
5	43	55.1	545	4	US-09-252-991A-30417	Sequence 30417, A	Sequence 30417, A	
6	42.5	54.5	1107	4	US-09-489-039A-8890	Sequence 8890, Ap	Sequence 8890, Ap	
7	42	53.8	30	4	US-09-205-258-1167	Sequence 1167, Ap	Sequence 1167, Ap	
8	42	53.8	419	4	US-09-252-991A-17499	Sequence 17499, A	Sequence 17499, A	
9	41	52.6	493	4	US-09-252-991A-20531	Sequence 20531, A	Sequence 20531, A	
10	41	52.6	511	4	US-09-252-991A-18691	Sequence 18691, A	Sequence 18691, A	
11	41	52.6	536	4	US-09-252-991A-22588	Sequence 22588, A	Sequence 22588, A	
12	41	52.6	595	4	US-09-252-991A-17469	Sequence 17469, A	Sequence 17469, A	
13	40	51.3	303	4	US-09-252-991A-22757	Sequence 22757, A	Sequence 22757, A	
14	40	51.3	355	4	US-09-252-991A-27426	Sequence 27426, A	Sequence 27426, A	
15	40	51.3	362	4	US-09-252-991A-30626	Sequence 30626, A	Sequence 30626, A	
16	40	51.3	419	4	US-09-252-991A-28487	Sequence 28487, A	Sequence 28487, A	
17	40	51.3	422	4	US-09-252-991A-18660	Sequence 18660, A	Sequence 18660, A	
18	40	51.3	518	4	US-09-252-991A-25967	Sequence 25967, A	Sequence 25967, A	
19	40	51.3	939	4	US-09-540-236-2399	Sequence 2399, Ap	Sequence 2399, Ap	
20	39	50.0	232	4	US-09-489-039A-10126	Sequence 10126, A	Sequence 10126, A	
21	39	50.0	542	1	US-08-412-431-3	Sequence 3, Appli	Sequence 3, Appli	
22	39	50.0	542	1	US-08-623-679-3	Sequence 3, Appli	Sequence 3, Appli	
23	39	50.0	542	3	US-08-933-774-3	Sequence 3, Appli	Sequence 3, Appli	
24	39	50.0	542	3	US-09-181-030-3	Sequence 3, Appli	Sequence 3, Appli	
25	39	50.0	542	4	US-09-534-242-3	Sequence 3, Appli	Sequence 3, Appli	
26	39	50.0	542	4	US-09-454-854-3	Sequence 3, Appli	Sequence 3, Appli	
27	39	50.0	542	4	US-09-164-671-3	Sequence 3, Appli	Sequence 3, Appli	

## ALIGNMENTS

## RESULT 1

```

US-09-252-991A-28695
; Sequence 28695: Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND
; TITLE OF INVENTION: AERUGINOSA FOR DIA
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,9
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,78
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,19
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28695
; LENGTH: 433
; TYPE: prt
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28695

```

Query Match 57.7%; Score 45; DB 4; Length 433;  
Best Local Similarity 64.3%; Pred. No. 4.2;  
Matches 9: Conservative 1: Mismatches 4: Indels

Qy 1 RQLPKRMGGSYRC 14  
| : | | | |  
Db 251 RRRLPVRPGGPGRC 264

## RESULT 2

```

RES001  2
US-09-252-991A-20615
Sequence 20615, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND A
TITLE OF INVENTION: AERUGINOSA FOR DIA
FILE REFERENCE: 107196 136
CURRENT APPLICATION NUMBER: US/09/252,9
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,78
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,19
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 20615
LENGTH: 164
TYPE: PRT

```

! ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20615

Query Match 55.1%; Score 43; DB 4; Length 164;  
Best Local Similarity 50.0%; Pred. No. 3.4;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 RORLPKRMGGSYRC 14  
|:|:|:|:|:|:|:  
Db 73 RRRMPRRSGPATRC 86

## RESULT 3

US-09-252-991A-30154  
; Sequence 30154, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:

! APPLICANT: Marc J. Rubenfield et al.  
! TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
! FILE REFERENCE: 107196.136  
! CURRENT APPLICATION NUMBER: US/09/252,991A  
! CURRENT FILING DATE: 1999-02-18  
! PRIOR APPLICATION NUMBER: US 60/074,788  
! PRIOR FILING DATE: 1998-02-18  
! PRIOR APPLICATION NUMBER: US 60/094,190  
! PRIOR FILING DATE: 1998-07-27  
! NUMBER OF SEQ ID NOS: 33142  
! SEQ ID NO 30154  
! LENGTH: 164  
! TYPE: PRT  
! ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30154

Query Match 55.1%; Score 43; DB 4; Length 164;  
Best Local Similarity 50.0%; Pred. No. 3.4;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 RORLPKRMGGSYRC 14  
|:|:|:|:|:|:|:  
Db 73 RRRMPRRSGPATRC 86

## RESULT 4

US-09-252-991A-18819  
; Sequence 18819, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:

! APPLICANT: Marc J. Rubenfield et al.  
! TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
! FILE REFERENCE: 107196.136  
! CURRENT APPLICATION NUMBER: US/09/252,991A  
! CURRENT FILING DATE: 1999-02-18  
! PRIOR APPLICATION NUMBER: US 60/074,788  
! PRIOR FILING DATE: 1998-02-18  
! PRIOR APPLICATION NUMBER: US 60/094,190  
! PRIOR FILING DATE: 1998-07-27  
! NUMBER OF SEQ ID NOS: 33142  
! SEQ ID NO 18819  
! LENGTH: 223  
! TYPE: PRT  
! ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18819

Query Match 55.1%; Score 43; DB 4; Length 223;  
Best Local Similarity 58.3%; Pred. No. 4.7;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 RORLPKRMGGSY 12  
|:|:|:|:|:|:|:  
Db 183 RQIRPRRRGGRRH 194

## RESULT 5

US-09-252-991A-30417  
; Sequence 30417, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:

! APPLICANT: Marc J. Rubenfield et al.  
! TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
! FILE REFERENCE: 107196.136  
! CURRENT APPLICATION NUMBER: US/09/252,991A  
! CURRENT FILING DATE: 1999-02-18  
! PRIOR APPLICATION NUMBER: US 60/074,788  
! PRIOR FILING DATE: 1998-02-18  
! PRIOR APPLICATION NUMBER: US 60/094,190  
! PRIOR FILING DATE: 1998-07-27  
! NUMBER OF SEQ ID NOS: 33142  
! SEQ ID NO 30417  
! LENGTH: 545  
! TYPE: PRT  
! ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30417

Query Match 55.1%; Score 43; DB 4; Length 545;  
Best Local Similarity 61.5%; Pred. No. 12;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RORLPKRMGGSYR 13  
|:|:|:|:|:|:|:  
Db 56 RSRLPRRAGGEVR 68

## RESULT 6

US-09-489-039A-8890  
; Sequence 8890, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:

! APPLICANT: Gary Breton et. al  
! TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
! FILE REFERENCE: 2709.2004001  
! CURRENT APPLICATION NUMBER: US/09/489,039A  
! CURRENT FILING DATE: 2000-01-27  
! PRIOR APPLICATION NUMBER: US 60/117,747  
! PRIOR FILING DATE: 1999-01-29  
! NUMBER OF SEQ ID NOS: 14342  
! SEQ ID NO 8890  
! LENGTH: 1107  
! TYPE: PRT  
! ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-8890

Query Match 54.5%; Score 42.5; DB 4; Length 1107;  
Best Local Similarity 64.3%; Pred. No. 33;  
Matches 9; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 RQR-LPKRMGGSYR 13  
|:|:|:|:|:|:|:  
Db 19 RQRVPDRLLGGQYR 32

## RESULT 7

US-09-205-258-1167  
; Sequence 1167, Application US/09205258  
; Patent No. 6525174  
; GENERAL INFORMATION:

! APPLICANT: Young et al.  
! TITLE OF INVENTION: 207 Human Secreted Proteins  
! FILE REFERENCE: P2007P1  
! CURRENT APPLICATION NUMBER: US/09/205,258  
! CURRENT FILING DATE: 1998-12-04  
! EARLIER APPLICATION NUMBER: PCT/US98/11422  
! EARLIER FILING DATE: 1998-06-04



FEATURE:  
 NAME/KEY: UNSURE  
 LOCATION: (313)  
 OTHER INFORMATION: Identity of amino acid at the above locations are unknown.  
 US-09-252-991A-20531

Query Match 52.6%; Score 41; DB 4; Length 492;  
 Best Local Similarity 50.0%; Pred. No. 25;  
 Matches 8; Conservative 3; Mismatches 1; Indels 4; Gaps 1;

Qy 3 RLPKRMGGSY----RC 14  
 |||:|:|:|  
 Db 361 RLPRRLGGDHPVPVC 376

RESULT 10  
 US-09-252-991A-18691  
 ; Sequence 18691, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 18691  
 ; LENGTH: 511  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-18691

Query Match 52.6%; Score 41; DB 4; Length 511;  
 Best Local Similarity 63.6%; Pred. No. 26;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 QRLPKRMGGSY 12  
 :|:|:|:|:|  
 Db 492 KRLPKKSGGSH 502

RESULT 11  
 US-09-252-991A-22588  
 ; Sequence 22588, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 22588  
 ; LENGTH: 556  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-22588

Query Match 52.6%; Score 41; DB 4; Length 556;  
 Best Local Similarity 63.6%; Pred. No. 29;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 RLPKRMGGSYR 13

Db 420 RLPRRLGGGQR 430  
 |||:|:|:|

RESULT 12  
 US-09-252-991A-17469  
 ; Sequence 17469, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 17469  
 ; LENGTH: 595  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-17469

Query Match 52.6%; Score 41; DB 4; Length 595;  
 Best Local Similarity 70.0%; Pred. No. 31;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 5 PKRMGGSYRC 14  
 |:|:|:|:|  
 Db 38 PRSGGSARC 47

RESULT 13  
 US-09-252-991A-22757  
 ; Sequence 22757, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 22757  
 ; LENGTH: 303  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-22757

Query Match 51.3%; Score 40; DB 4; Length 303;  
 Best Local Similarity 57.1%; Pred. No. 23;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGGSYRC 14  
 |||:|:|:|  
 Db 61 RQRLRRWSGRRC 74

RESULT 14  
 US-09-252-991A-27426  
 ; Sequence 27426, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 27426

; LENGTH: 355

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-27426

Query Match

Best Local Similarity 51.3%; Score 40; DB 4; Length 355;

Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 RORLPKRMGGSY 12

Db 103 RRVDPRLGGAH 114

RESULT 15

US-09-252-991A-30626

; Sequence 30626, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 30626

; LENGTH: 362

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30626

Query Match

Best Local Similarity 51.3%; Score 40; DB 4; Length 362;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RORLPKRMGG 10

Db 70 RRLPRRYGG 79

Search completed: June 3, 2004, 14:11:34

Job time : 7.525 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 14:11:58 ; Search time 19.075 Seconds  
(without alignments)  
206.487 Million cell updates/sec

Title: US-09-745-008-14  
Perfect score: 78  
Sequence: 1 RQLPKRMGGSYRC 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	14	9	US-09-745-008-14
2	78	100.0	21	9	US-09-745-008-13
3	78	100.0	1162	9	US-09-745-008-34
4	67.5	86.5	45	9	US-09-745-008-12
5	43	55.1	142	12	US-10-425-114-58669
6	43	55.1	155	12	US-10-425-114-51601
7	43	55.1	162	12	US-10-425-114-47458
8	43	55.1	164	12	US-10-425-114-39475
9	43	55.1	165	12	US-10-425-114-38751
10	43	55.1	165	12	US-10-425-114-44961
11	43	55.1	166	12	US-10-425-114-39600
12	43	55.1	166	12	US-10-425-114-60911
13	43	55.1	166	12	US-10-425-114-62444
14	43	55.1	169	12	US-10-425-114-67338
15	43	55.1	179	12	US-10-425-114-61006

16	43	55.1	179	12	US-10-425-114-62787	Sequence 62787, A
17	43	55.1	181	12	US-10-425-114-39542	Sequence 39542, A
18	43	55.1	181	12	US-10-425-114-72129	Sequence 72129, A
19	43	55.1	181	12	US-10-425-114-72143	Sequence 72143, A
20	43	55.1	183	12	US-10-425-114-58636	Sequence 58636, A
21	43	55.1	184	12	US-10-425-114-55429	Sequence 55429, A
22	43	55.1	185	12	US-10-425-114-47709	Sequence 47709, A
23	43	55.1	185	12	US-10-425-114-65756	Sequence 65756, A
24	42	53.8	30	10	US-09-933-767-1167	Sequence 1167, Ap
25	42	53.8	30	12	US-10-004-860-1167	Sequence 1167, Ap
26	42	53.8	30	14	US-10-023-282-1167	Sequence 1167, Ap
27	41	52.6	564	15	US-10-369-493-585	Sequence 585, App
28	40.5	51.9	235	14	US-10-240-154-22	Sequence 22, Appl
29	40	51.3	60	12	US-10-424-599-176730	Sequence 176730, A
30	40	51.3	147	12	US-10-424-599-152415	Sequence 152415, A
31	40	51.3	150	14	US-10-156-761-10416	Sequence 10416, A
32	40	51.3	185	12	US-10-425-114-60116	Sequence 60116, A
33	40	51.3	230	9	US-09-741-669-372	Sequence 372, App
34	40	51.3	230	12	US-10-282-122A-42769	Sequence 42769, A
35	40	51.3	259	15	US-10-104-047-2303	Sequence 2303, Ap
36	40	51.3	285	12	US-10-425-114-49272	Sequence 49272, A
37	40	51.3	302	15	US-10-094-886-40	Sequence 40, Appl
38	40	51.3	302	15	US-10-094-886-42	Sequence 42, Appl
39	40	51.3	533	15	US-10-310-154-699	Sequence 699, App
40	40	51.3	559	15	US-10-094-749-1691	Sequence 1691, Ap
41	40	51.3	869	16	US-10-471-115-18	Sequence 18, Appl
42	40	51.3	1315	15	US-10-094-886-52	Sequence 52, Appl
43	40	51.3	1386	15	US-10-094-886-38	Sequence 38, Appl
44	39	50.0	153	12	US-10-425-114-68203	Sequence 68203, A
45	39	50.0	176	12	US-10-425-114-72617	Sequence 72617, A

#### ALIGNMENTS

#### RESULT 1

US-09-745-008-14  
; Sequence 14, Application US/09745008  
; Patent No. US20020137667A1  
; GENERAL INFORMATION:  
; APPLICANT: Chuenkova, Marina  
; APPLICANT: Pereira, Marcio A.  
; TITLE OF INVENTION: T. Cruz-Deived Neurotrophic Agents and  
; TITLE OF INVENTION: Methods of Use Therefor  
; FILE REFERENCE: 1322.1028-001  
; CURRENT APPLICATION NUMBER: US/09/745,008  
; CURRENT FILING DATE: 2000-12-20  
; PRIOR FILING DATE: 1999-12-20  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic peptide  
US-09-745-008-14

Query Match 100.0%; Score 78; DB 9; Length 14;  
Best Local Similarity 100.0%; Pred. No. 2e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQLPKRMGGSYRC 14

Db 1 RQLPKRMGGSYRC 14

#### RESULT 2

US-09-745-008-13  
; Sequence 13, Application US/09745008  
; Patent No. US20020137667A1  
; GENERAL INFORMATION:

; APPLICANT: Chuenkova, Marina  
; APPLICANT: Pereira, Mircio A.  
; TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and  
; TITLE OF INVENTION: Methods of Use Therefor  
; FILE REFERENCE: 1322.1028-001  
; CURRENT APPLICATION NUMBER: US/09/745,008  
; CURRENT FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: US 60/172,881  
; PRIOR FILING DATE: 1999-12-20  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic peptide  
US-09-745-008-13

Query Match 100.0%; Score 78; DB 9; Length 21;  
Best Local Similarity 100.0%; Pred. No. 3e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGGSYRC 14  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 RQRLPKRMGGSYRC 14

RESULT 3  
US-09-745-008-34

; Sequence 34, Application US/09745008  
; Patent No. US20020137667A1  
; GENERAL INFORMATION:  
; APPLICANT: Chuenkova, Marina  
; APPLICANT: Pereira, Mircio A.  
; TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and  
; TITLE OF INVENTION: Methods of Use Therefor  
; FILE REFERENCE: 1322.1028-001  
; CURRENT APPLICATION NUMBER: US/09/745,008  
; CURRENT FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: US 60/172,881  
; PRIOR FILING DATE: 1999-12-20  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 34  
; LENGTH: 1162  
; TYPE: PRT  
; ORGANISM: Trypanosoma cruzi  
US-09-745-008-34

Query Match 100.0%; Score 78; DB 9; Length 1162;  
Best Local Similarity 100.0%; Pred. No. 0.00016;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGGSYRC 14  
| | | | | | | | | | | | | | | | | | | | | |  
Db 379 RQRLPKRMGGSYRC 392

RESULT 4  
US-09-745-008-12

; Sequence 12, Application US/09745008  
; Patent No. US20020137667A1  
; GENERAL INFORMATION:  
; APPLICANT: Chuenkova, Marina  
; APPLICANT: Pereira, Mircio A.  
; TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and  
; TITLE OF INVENTION: Methods of Use Therefor  
; FILE REFERENCE: 1322.1028-001  
; CURRENT APPLICATION NUMBER: US/09/745,008  
; CURRENT FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: US 60/172,881  
; PRIOR FILING DATE: 1999-12-20

; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 45  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic peptide  
US-09-745-008-12

Query Match 86.5%; Score 67.5; DB 9; Length 45;  
Best Local Similarity 93.3%; Pred. No. 0.00039;  
Matches 14; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 RQRLPKRMGGSYRC 14  
| | | | | | | | | | | | | | | | | | | | | |  
Db 24 RQRLPKRMGGSYRC 38

RESULT 5

US-10-425-114-58669  
; Sequence 58669, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 58669  
; LENGTH: 142  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700615246\_F11.pep  
US-10-425-114-58669

Query Match 55.1%; Score 43; DB 12; Length 142;  
Best Local Similarity 70.0%; Pred. No. 18;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGG 10  
| | | | | | | | | | | | | | | | | | | | | |  
Db 108 RRRLPRELGG 117

RESULT 6

US-10-425-114-51601  
; Sequence 51601, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 51601  
; LENGTH: 155  
; TYPE: PRT

```
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700222827_FLI.pep
US-10-425-114-51601

Query Match      55.1%; Score 43; DB 12; Length 155;
Best Local Similarity 70.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RQRLPKRMGG 10
Db      133 RRLPRRLGG 142
      |:|:|:|:|:|

RESULT 7
US-10-425-114-47458
; Sequence 47458, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 47458
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700106619_FLI.pep
US-10-425-114-47458

Query Match      55.1%; Score 43; DB 12; Length 162;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 RQRLPKRMGG 10
Db      133 RRLPRRLGG 142
      |:|:|:|:|:|

RESULT 8
US-10-425-114-39475
; Sequence 39475, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 39475
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700207896_FLI.pep
US-10-425-114-39475
```

```
Query Match      55.1%; Score 43; DB 12; Length 164;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 RQRLPKRMGG 10
Db      108 RRLPRRLGG 117
      |:|:|:|:|:|

RESULT 9
US-10-425-114-38751
; Sequence 38751, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 38751
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73069B07_FLI.pep
US-10-425-114-38751

Query Match      55.1%; Score 43; DB 12; Length 165;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 RQRLPKRMGG 10
Db      108 RRLPRRLGG 117
      |:|:|:|:|:|

RESULT 10
US-10-425-114-44961
; Sequence 44961, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 44961
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700448317_FLI.pep
US-10-425-114-44961

Query Match      55.1%; Score 43; DB 12; Length 165;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 RQRLPKRMGG 10
```



; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 61006  
; LENGTH: 179  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3279-001-D10\_FLI.pep  
US-10-425-114-61006

Query Match 55.1%; Score 43; DB 12; Length 179;  
Best Local Similarity 70.0%; Pred. No. 23;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGG 10  
|:|:|:|:  
Db 133 RRLPRRLGG 142

Search completed: June 3, 2004, 14:23:13  
Job time : 20.075 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 14:13:28 ; Search time 4.025 Seconds  
(without alignments)  
181.114 Million cell updates/sec

Title: US-09-745-008-14  
Perfect score: 78  
Sequence: 1 RQRLPKRMGGSYRC 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	1162	1 TCNA TRYCR	P23253 trypanosoma
2	42	53.8	1169	1 IF2 PROMP	Q7u229 prochloroco
3	40.5	51.9	235	1 TNFA RAT	P16599 rattus norv
4	40	51.3	210	1 YQEH ECOLI	Q46941 escherichia
5	40	51.3	428	1 YL28 LACLA	Q9cdd3 lactococcus
6	39	50.0	304	1 K2S3 HUMAN	Q14952 homo sapien
7	39	50.0	311	1 STCC EMENI	Q00668 emericella
8	39	50.0	520	1 IBMP CAMVP	P18617 cauliflowe
9	39	50.0	1319	1 MN1 HUMAN	Q10571 homo sapien
10	38	48.7	361	1 IDH1 KLULA	Q94229 kluyveromyc
11	38	48.7	414	1 SX17 HUMAN	Q9h612 homo sapien
12	38	48.7	553	1 SCAP CHICK	Q14410 homo sapien
13	38	48.7	637	1 SCAP CHICK	Q92075 gallus gall
14	38	48.7	1403	1 PROS DROME	P29617 drosophila
15	38	48.7	1556	1 PROS DROVI	Q9uea1 drosophila
16	37.5	48.1	54	1 RS14 AERPE	P58731 aeropyrum p
17	37	47.4	197	1 IE68 HSV2	P14379 herpes simp
18	37	47.4	253	1 YDEO ECOLI	P76135 escherichia
19	37	47.4	304	1 K2S1 HUMAN	Q14954 homo sapien
20	37	47.4	304	1 K2S2 HUMAN	P43631 homo sapien
21	37	47.4	304	1 K2S4 HUMAN	P43632 homo sapien
22	37	47.4	304	1 K2S5 HUMAN	Q14953 homo sapien
23	37	47.4	341	1 K2L3 HUMAN	P43628 h killer ce
24	37	47.4	348	1 K2L1 HUMAN	P43626 h killer ce
25	37	47.4	348	1 K2L2 HUMAN	P43627 homo sapien
26	37	47.4	387	1 K3S1 HUMAN	Q14943 homo sapien
27	37	47.4	444	1 K3L1 HUMAN	P43629 h killer ce
28	37	47.4	455	1 K3L2 HUMAN	P43630 homo sapien
29	37	47.4	504	1 MEPA SOLTU	P29677 solanum tub
30	37	47.4	520	1 IBMP CAMVS	P03559 cauliflowe
31	37	47.4	646	1 MU18 HUMAN	P43121 homo sapien
32	36.5	46.8	234	1 TNFA CAVPO	P51435 cavia porce
33	36.5	46.8	235	1 TNFA_MOUSE	P06804 mus musculu

34	36.5	46.8	2032	1 TRPG CAEEL	Q93971 caenorhabdi
35	36	46.2	87	1 RL34 SULTO	Q975k6 sulfobolus
36	36	46.2	89	1 RL34 METJA	P54053 methanococc
37	36	46.2	97	1 SY07 MOUSE	Q03366 mus musculu
38	36	46.2	97	1 SY07 RAT	Q9qkx8 rattus norv
39	36	46.2	192	1 VMTZ LAMBD	P03731 bacterioph
40	36	46.2	196	1 RL11 CAEEL	Q94300 caenorhabdi
41	36	46.2	221	1 FLXW RHILE	P14332 rhizobium l
42	36	46.2	255	1 YORE PSECL	Q03003 pseudomonas
43	36	46.2	257	1 RNH2 BACCR	Q819w9 bacillus ce
44	36	46.2	366	1 ALR XANAC	Q8pgd0 xanthomonas
45	36	46.2	452	1 MTN3 CHICK	O42401 gallus gall

## ALIGNMENTS

## RESULT 1

ID	TCNA_TRYCR	STANDARD;	PRT;	1162 AA.
AC	P23253;			
DT	01-NOV-1991 (Rel. 20, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Sialidase (EC 3.2.1.18) (Neuraminidase) (NA) (Major surface antigen).			
GN	TCNA.			
OS	Trypanosoma cruzi.			
OC	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.			
OX	NCBI_TaxID=5693;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Silvio X-10/4;			
RX	MEDLINE=91277609; PubMed=1711561;			
RA	Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D., Prioli R.P.;			
RT	"The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial neuraminidases, YMTD repeats of the low density lipoprotein receptor, and type III modules of fibronectin.";			
RL	J. Exp. Med. 174:179-191(1991).			
RN	[2]			
RP	SUBCELLULAR LOCATION.			
RX	MEDLINE=91376547; PubMed=1896773;			
RA	Prioli R.P., Mejia J.S., Aji T., Aikawa M., Pereira M.E.A.;			
RT	Trypanosoma cruzi: localization of neuraminidase on the surface of trypomastigotes.";			
RL	trop. Med. Parasitol. 42:146-150(1991).			
CC	- FUNCTION: Developmentally regulated neuraminidase implicated in parasite invasion of cells.			
CC	- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2-3)-, alpha-(2-6)-, alpha-(2-8)-glycosidic linkages of terminal sialic residues in oligosaccharides, glycoproteins, glycolipids, colominic acid and synthetic substrates.			
CC	- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (Potential).			
CC	- DEVELOPMENTAL STAGE: Maximal activity in trypomastigotes, minimum in epimastigotes and not detectable in amastigotes.			
CC	- MISCELLANEOUS: The variable lengths of the long tandem repeat domain could account in part for the polymorphism of the TCNA protein.			
CC	- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.			
CC	- SIMILARITY: Contains 3 ENR repeats.			

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; M61732; AAA30255.1; -

PIR; JH0557; JH0557.

InterPro; IPR002860; GH\_BNR.

```
DR InterPro: IPR008377; Sialidase_trypan.
DR Pfam: PF02012; ENR; 2.
DR PRINTS: PR01803; TCSIALIDASE.
KW Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
FT Phosphorylation.
KW DOMAIN 1 457 CYS-RICH.
FT REPEAT 23 34 ENR 1.
FT REPEAT 163 174 ENR 2.
FT REPEAT 209 220 ENR 3.
FT DOMAIN 458 588
FT DOMAIN 589 1120
FT CARBOHYD 342 342 44 X 12 AA TANDEM REPEATS, LTR DOMAIN.
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1162 AA; 07049221897C6A40 CRC64;

Query Match 100.0%; Score 78; DB 1; Length 1162;
Best Local Similarity 100.0%; Pred. No. 4.2e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RQRLPKRMGGSYRC 14
|||||
DB 379 RQRLPKRMGGSYRC 392

RESULT 2
IF2_PROMP
ID IF2_PROMP STANDARD; PRT; 1169 AA.
AC Q7U2Z9;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Translation initiation factor IF-2.
GN INFS OR PMW1494.
OS Prochlorococcus marinus subsp. pastoris (strain COMP 1378 / MED4).
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=59919;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825698; PubMed=12917642;
RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
RA Algren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA Webb E.A., Zinser E.R., Chisholm S.W.;
RA "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RA niche differentiation.";
RL Nature 424:1042-1047(2003).
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the IF-2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; BX572094; CAEL19953.1; -.
CC HAWAP; MF_00100; -.
CC InterPro: IPR004161; EFTU_D2.
CC InterPro: IPR000795; EF_GTPbind.
CC InterPro: IPR000178; IF2.
CC InterPro: IPR006847; IF2_N.
CC InterPro: IPR005225; Small_GTP.
CC Pfam; PF00009; GTP_EFTU; 1.

DR InterPro: IPR008377; Sialidase_trypan.
DR Pfam: PF04760; IF2_N; 2.
DR ProDom; PD186100; IF2; 1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; 1.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome. 816 G-DOMAIN.
FT DOMAIN 664 677 GTP (BY SIMILARITY).
FT NP_BIND 720 724 GTP (BY SIMILARITY).
FT NP_BIND 774 777 GTP (BY SIMILARITY).
SQ SEQUENCE 1169 AA; 127147 MW; 0705D909F5377562 CRC64;

Query Match 53.8%; Score 42; DB 1; Length 1169;
Best Local Similarity 61.5%; Pred. No. 14;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RQRLPKRMGGSYR 13
|||:|:|:|
DB 292 RQGVNREGGYPYR 304

RESULT 3
TNFA_RAT
ID TNFA_RAT STANDARD; PRT; 235 AA.
AC P16599; Q9J126; Q9J127;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tumor necrosis factor precursor (TNF-alpha) (Tumor necrosis factor
DE ligand superfamily member 2) (TNF-a) (Cachectin).
GN TNF OR TNFSE2 OR TNFA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Shirai T., Shimizu N., Horiguchi S., Ito H.;
RT "Cloning and expression in Escherichia coli of the gene for rat tumor
RT necrosis factor.";
RL Agric. Biol. Chem. 53:1733-1736(1989).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92329007; PubMed=1627266;
RA Estler H.C., Grewe M., Gaussling R., Pavlovic M., Decker K.;
RA "Rat tumor necrosis factor-alpha. Transcription in rat Kupffer cells
RA and in vitro posttranslational processing based on a PCR-derived
RA cDNA.";
RL Biol. Chem. Hoppe-Seyler 373:271-281(1992).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=94040766; PubMed=8224868;
RA Kwon J., Chung I.Y., Benveniste E.N.;
RA "Cloning and sequence analysis of the rat tumor necrosis
RA factor-encoding genes.";
RL Gene 132:227-236(1993).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=ACI/SegHsd, BB(DR)/Wor, BN/SeNHsd, DA/Bk1, F344/NHsd, and
RC LEW/NHsd;
RX MEDLINE=21369712; PubMed=11477479;
RA Furuya T., Joe B., Salstrom J.L., Hashiramoto A., Dobbins D.E.,
RA Wilder R.L., Remmers E.F.;
RT "Polymorphisms of the tumor necrosis factor alpha locus among
RT autoimmune disease susceptible and resistant inbred rat strains.";
RN Genes Immun. 2:229-232(2001).
[5]
RP SEQUENCE FROM N.A.
RA Decker K.F.;
RA Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
```



[6]  
 RP SEQUENCE FROM N.A., AND VARIANTS PRO-122 AND GLU-190.  
 RC STRAIN=Dark Agouti;  
 RA Seidel M.F., Junier M.-P., Vetter H.;  
 RT "TNF-alpha polymorphism in rats with collagen-induced arthritis.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 [7]  
 RN SEQUENCE OF 1-231 FROM N.A.  
 RP TISSUE=tail;  
 RC Kirieits M.J., Vardimon D., Kunz H.W., Gill T.J. III;  
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Cytokine that binds to TNFRSF1A/TNFR1 and  
 CC TNFRSF1B/TNFR. It is mainly secreted by macrophages and can  
 CC induce cell death of certain tumor cell lines. It is potent  
 CC pyrogen causing fever by direct action or by stimulation of  
 CC interleukin 1 secretion and is implicated in the induction of  
 CC cachexia. Under certain conditions it can stimulate cell  
 CC proliferation and induce cell differentiation.  
 CC -1- SUBUNIT: Homotrimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists as an  
 CC extracellular soluble form (By similarity).  
 CC -1- PTM: The soluble form derives from the membrane form by  
 CC proteolytic processing (By similarity).  
 CC -1- PTM: The membrane form, but not the soluble form, is  
 CC phosphorylated on serine residues. Dephosphorylation of the  
 CC membrane form occurs by binding to soluble TNFRSF1A/TNFR1 (By  
 CC similarity).  
 CC -1- DISEASE: Cachexia accompanies a variety of diseases, including  
 CC cancer and infection, and is characterized by general ill health  
 CC and malnutrition.  
 CC -1- SIMILARITY: Belongs to the tumor necrosis factor family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D00475; EAA00367.1; -  
 CC EMBL; X66539; CAA47146.1; -  
 CC EMBL; L00981; AAL16275.1; -  
 CC EMBL; AF329982; AAK53568.1; -  
 CC EMBL; AF329983; AAK53569.1; -  
 CC EMBL; AF329984; AAK53570.1; -  
 CC EMBL; AF329985; AAK53571.1; -  
 CC EMBL; AF329986; AAK53572.1; -  
 CC EMBL; AF329987; AAK53573.1; -  
 CC EMBL; AJ002278; CAA05290.1; -  
 CC EMBL; AF269159; AAF82567.1; -  
 CC EMBL; AF269160; AAF82568.1; -  
 CC EMBL; L19123; AAA42255.1; -  
 CC PIR; JU0029; JU0029.  
 CC HSP; P06804; 2TNF.  
 CC InterPro; IPR006053; TNF abc.  
 CC InterPro; IPR006052; TNF\_family.  
 CC InterPro; IPR008983; TNF\_like.  
 CC InterPro; IPR003636; TNF\_subf.  
 CC Pfam; PF00229; TNF; 1.  
 CC PRINTS; PR01234; TNECROSISFCT.  
 CC ProDom; PD002012; TNF\_subf; 1.  
 CC SMART; SM00207; TNF; 1.  
 CC PROSITE; PS00251; TNF; 1.  
 CC PROSITE; PS00049; TNF; 2; 1.  
 CC Cytokine; Transmembrane; Signal-anchor; Phosphorylation.  
 CC CHAIN 1 235 TUMOR NECROSIS FACTOR, MEMBRANE FORM.  
 CC CYTOKINE; TUMOR NECROSIS FACTOR, SOLUBLE FORM.  
 CC DOMAIN 80 235 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 36 56 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC DOMAIN 57 235 EXTRACELLULAR (POTENTIAL).  
 CC SITE 79 80 CLEAVAGE (BY ADAM17) (BY SIMILARITY).  
 CC

FT MOD RES 2 2 PHOSPHORYLATION (BY CKI) (BY SIMILARITY).  
 FT DISULFID 148 179 BY SIMILARITY.  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 122 122 L -> P.  
 FT VARIANT 190 190 K -> E.  
 FT CONFLICT 39 39 L -> P (IN REF. 2 AND 5).  
 FT CONFLICT 163 163 I -> T (IN REF. 2 AND 5).  
 FT CONFLICT 202 202 F -> S (IN REF. 2 AND 5).  
 SQ SEQUENCE 235 AA; 25806 MW; B808EC6D049C2F3B CRC64;  
 Query Match 51.9%; Score 40.5; DB 1; Length 235;  
 Best Local Similarity 56.2%; Pred. No. 5.2;  
 Matches 9; Conservative 2; Mismatches 2; Indels 3; Gaps 1;  
 QY 2 QRLPKRMGG---SYRC 14  
 : |||: |||  
 Db 15 EALPKRMGGGLQNSRC 30  
 : |||: |||  
 RESULT 4  
 YQEH\_ECOLI STANDARD; PRT; 210 AA.  
 ID YQEH\_ECOLI  
 AC Q46941;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein yqeh.  
 GN YQEH OR B2846.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MGL1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474 (1997).  
 CC -1- SIMILARITY: SOME, TO E.COLI YKKG.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U29581; AAB40493.1; ALT\_INIT.  
 CC EMBL; AE003368; AAC75885.1; ALT\_INIT.  
 CC EcoGene; EGI30399; Yqeh.  
 CC InterPro; IPR000792; HTH LuxR.  
 CC Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 210 AA; 24287 MW; 9C0778FECE9823C CRC64;  
 Query Match 51.3%; Score 40; DB 1; Length 210;  
 Best Local Similarity 70.0%; Pred. No. 5.7;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 QRLPKRMGG 10  
 : |||: |||  
 Db 187 QRIEKRMMGG 196  
 : |||: |||  
 RESULT 5  
 YL28\_LACLA STANDARD; PRT; 428 AA.  
 ID YL28\_LACLA  
 AC Q9CDT3;  
 DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical zinc metalloprotease LL2128 (EC 3.4.24.-).  
 GN LL2128.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 CC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
 OX NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403.  
 RX MEDLINE=21235186; PubMed=11337471;  
 RA Boletín A., Wincker P., Mauger S., Jaillon O., Malarne K.,  
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RT lactis ssp. lactis IL1403";  
 RL Genome Res. 11:731-753(2001).  
 CC -!- COPACTOR: Zinc (Probable).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: Belongs to peptidase family M50B.  
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AE006441; AAK06226.1; -.  
 DR F1R; H86890; H86890.  
 DR MEROPS; M50.004; -.  
 DR InterPro; IPR001478; PDZ.  
 DR InterPro; IPR004387; Pept\_M50\_Zn.  
 DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 DR InterPro; IPR008915; Peptidase\_M50.  
 DR Pfam; PF00595; PDZ; 1.  
 DR Pfam; PF02163; Peptidase\_M50; 1.  
 DR SMART; SM00228; PDZ; 1.  
 DR TIGRFAMs; TIGR00054; TIGR00054; 1.  
 DR PROSITE; PS50106; PDZ; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;  
 KW Complete proteome.  
 FT METAL 19 19 ZINC (CATALYTIC) (POTENTIAL).  
 FT ACT SITE 20 20 POTENTIAL.  
 FT METAL 23 23 ZINC (CATALYTIC) (POTENTIAL).  
 FT TRANSMEM 188 210 POTENTIAL.  
 FT TRANSMEM 354 376 POTENTIAL.  
 FT TRANSMEM 401 423 POTENTIAL.  
 FT DOMAIN 188 282 PDZ.  
 SQ SEQUENCE 428 AA; 46423 MW; 5BE5447331D56AC8 CRC64;  
 Query Match 51.3%; Score 40; DB 1; Length 428;  
 Best Local Similarity 77.8%; Pred. No. 12;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 5 PKRMGGSYR 13  
 Db 280 PKRMGGSYR 288  
 ||:|||||  
 RESULT 6  
 K2S3 HUMAN  
 ID K2S3 HUMAN STANDARD; PRT; 304 AA.  
 AC Q14952; Q00644;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Killer cell immunoglobulin-like receptor 2DS3 precursor (MHC class I  
 DE NK cell receptor) (Natural killer associated transcript 7) (NKAT-7).  
 GN KIR2DS3 OR NKAT7.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96270004; PubMed=8662091;  
 RA Doebering C., Samaridis J., Colonna M.;  
 RT "Alternatively spliced forms of human killer inhibitory receptors.";  
 RL Immunogenetics 44:227-230(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymphoid;  
 RA Biassoni R.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Receptor on natural killer (NK) cells for HLA-C alleles.  
 CC Does not inhibit the activity of NK cells.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily.  
 CC -!- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; L76670; AAB36598.1; -.  
 DR EMBL; X97231; CAA65870.1; -.  
 DR HSSP; P43626; LNKR.  
 DR Genew; HGNC:6335; KIR2DS3.  
 DR MIN; 604954; -.  
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
 DR GO; GO:0006968; P:cellular defense response; TAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003599; Ig.  
 DR Pfam; PF00047; Ig; 2.  
 DR SMART; SM00409; Ig; 1.  
 DR PROSITE; PS50835; IG LIKE; FALSE NEG.  
 DR Receptor; Immunoglobulin domain; Glycoprotein; Signal; Transmembrane;  
 KW Repeat; Multigene family.  
 FT SIGNAL 1 21 BY SIMILARITY.  
 FT CHAIN 22 304 KILLER CELL IMMUNOGLOBULIN-LIKE RECEPTOR  
 FT DSS3.  
 FT DOMAIN 22 245 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 246 264 POTENTIAL.  
 FT DOMAIN 265 304 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 42 107 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 142 205 IG-LIKE C2-TYPE 2.  
 FT DISULFID 49 100 BY SIMILARITY.  
 FT DISULFID 149 198 BY SIMILARITY.  
 FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 304 AA; 33717 MW; 80693F79844F9E7E CRC64;  
 Query Match 50.0%; Score 39; DB 1; Length 304;  
 Best Local Similarity 60.0%; Pred. No. 13;  
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 PKRMGGSYR 14  
 Db 189 PATQGGTYRC 198  
 ||:|||||  
 RESULT 7  
 STCC\_EMENI  
 ID STCC\_EMENI STANDARD; PRT; 311 AA.  
 AC Q00668;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Putative sterigmatocystin biosynthesis peroxidase stcC precursor  
DE (EC 1.11.1.-).  
OS STCC.  
GN Emericella nidulans (Aspergillus nidulans).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; Emericella.  
OX NCBI\_TaxID=162425;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FGSC 26;  
RX MEDLINE=96202293; PubMed=8643646;  
RA Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,  
RA Keller N.P., Adams T.H., Leonard T.J.;  
RT "Twenty-five coregulated transcripts define a sterigmatocystin gene  
RT cluster in *Aspergillus nidulans*."  
RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).  
CC -!- PATHWAY: Sterigmatocystin biosynthesis.  
CC -!- SIMILARITY: BELONGS TO THE CHLOROPEROXIDASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
CC  
CC -----  
DR EMBL; U34740; AAC49193.1; -.  
DR InterPro; IPR000028; Chloroperoxidase.  
DR Pfam; PF01328; Peroxidase 2; 1.  
DR ProDom; PD040763; Chloroperoxidase; 1.  
KW Oxidoreductase; Peroxidase; Iron; Signal.  
FT SIGNAL 1 25  
FT CHAIN 26 311  
FT METAL 44 44  
FT SEQUENCE 311 AA; 34816 MW; 18CB48B015CED735 CRC64;  
Query Match 50.0%; Score 39; DB 1; Length 311;  
Best Local Similarity 66.7%; Pred. No. 13;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 2 QRLPKRMGGSYR 13  
DB 215 ERLPVKGGGAR 226  
RESULT 8  
ID IBMP CAMVP STANDARD; PRT; 520 AA.  
AC F18617;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Inclusion body matrix protein (Viroplasmin).  
GN VI.  
OS Cauliflower mosaic virus (strain PV147) (CaMV).  
OC Viruses; Retroviruses; Caulimoviridae; Caulimovirus.  
OX NCBI\_TaxID=10647;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90384848; PubMed=2402462;  
RA Volovitch M., Modjtahedi N., Chouik Y., Yot P.;  
RT "DNA sequence of gene VI of cauliflower mosaic virus strain PV147."  
RL Nucleic Acids Res. 18:5297-5297(1990).  
CC -!- FUNCTION: ENHANCES THE TRANSLATION OF DOWNSTREAM ORF'S ON  
CC POLYCYSTRONIC MRNA'S DERIVED FROM CAULIFLOWER MOSAIC VIRUS.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic inclusion bodies.  
CC -!- MISCELLANEOUS: THE INCLUSION BODIES ARE THE SITE OF VIRAL DNA  
CC SYNTHESIS, VIRION ASSEMBLY AND ACCUMULATION IN THE INFECTED CELL.  
CC -!- SIMILARITY: Belongs to the caulimoviruses viroprotein family.  
CC  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
CC  
CC -----  
DR EMBL; X53860; CAA37853.1; -.  
DR FIR; S11217; S11217.  
DR InterPro; IPR002609; Caulimo VI.  
DR InterPro; IPR009027; L9\_N like.  
DR Pfam; PF01693; Caulimo VI; 1.  
KW Trans-acting factor; Translation regulation.  
SQ SEQUENCE 520 AA; 57746 MW; 4CB77C155BE61D65 CRC64;  
Query Match 50.0%; Score 39; DB 1; Length 520;  
Best Local Similarity 46.2%; Pred. No. 22;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 2 QRLPKRMGGSYR 14  
DB 464 ERTVEKGGSYK 476  
RESULT 9  
ID MN1\_HUMAN STANDARD; PRT; 1319 AA.  
AC Q10571;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Probable tumor suppressor protein MN1.  
GN MN1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=95249266; PubMed=7731706;  
RA Deprez R.H.I., Riegman P.H.J., Groen N.A., Warringa U.L.,  
RA van Biezen N.A., Molijn A.C., Bootsma D., de Jong P.J.,  
RA Menon A.G., Kley N.A., Seizenger B.R., Zwarthoff E.C.;  
RT "Cloning and characterization of MN1, a gene from chromosome 22q11,  
RT which is disrupted by a balanced translocation in a meningioma."  
RL Oncogene 10:1521-1528(1995).  
RN [2]  
RP SEQUENCE OF 1304-1319 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=97145634; PubMed=9026990;  
RA Dmitrenko V.V., Garifulin O.M., Shostak E.A., Smikodub A.I.,  
RA Kavsan V.M.;  
RT "The characteristics of different types of mRNA expressed in the human  
RT brain."  
RL Cyt. Genet. (Russ.) 30:41-47(1996).  
CC -!- FUNCTION: May play a role in tumor suppression.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q10571-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q10571-2; Sequence=Not described;  
CC Note=No experimental confirmation available;  
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed. Highest levels in  
CC skeletal muscle.  
CC -!- DISEASE: Involved in a form of acute myeloid leukemia (AML) by a  
CC chromosomal translocation t(12;22)(p13;q11) that involves MN1 and  
CC TEL.  
CC -!- DISEASE: Defects in MN1 may be a cause of meningiomas, slowly  
CC growing benign tumors derived from the arachnoidal cap cells of  
CC the leptomeninges, the soft coverings of the brain and spinal  
CC cord. Meningiomas are believed to be the most common primary

```
CC tumors of the central nervous system in man.
CC -!- CAUTION: It is uncertain whether Met-1 or Met-30 is the initiator.
CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.; URL="http://www.infobiogen.fr/services/chronocancer/Genes/MN1.html".
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/MN1.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X82209; CAA57693.1; ALT_INIT.
CC EMBL; Z70218; CAA94179.1; -.
CC Genew; HGNC:7180; MN1.
CC MIM; 156100; -.
CC MIM; 607174; -.
CC KW Anti-oncogene; Chromosomal translocation; Alternative splicing.
CC FT DOMAIN 295 309 POLY-GLN.
CC FT DOMAIN 523 550 POLY-GLN.
CC SQ SEQUENCE 1319 AA; 135943 MW; 21197C9BBA272BE2 CRC64;
CC -----
CC Query Match 50.0%; Score 39; DB 1; Length 1319;
CC Best Local Similarity 87.5%; Pred. No. 57;
CC Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 4 LPRRMGGS 11
CC Db 643 LPRRMGGS 650
CC -----
CC RESULT 10
CC IDH1_KLUJA
CC ID IDH1_KLUJA STANDARD; PRT; 361 AA.
CC AC 094223;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Isocitrate dehydrogenase [NAD] subunit 1, mitochondrial precursor
CC DE (EC 1.1.1.41) (Isocitric dehydrogenase) (NAD+-specific ICDH).
CC GN IDH1.
CC OS Kluyveromyces lactis (Yeast).
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
CC OX NCBI_TaxID=28985;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=JBD100;
CC RX MEDLINE=20426969; PubMed=10975257;
CC RA Eizinga S.D.J., van Oosterum K., Maat C., Grivell L.A.,
CC RA van der Spek H.;
CC RT "Isolation and RNA-binding analysis of NAD+-isocitrate
CC RT dehydrogenases from Kluyveromyces lactis and Schizosaccharomyces
CC RT pombe.";
CC RL Curr. Genet. 38:87-94(2000).
CC CC -!- FUNCTION: Performs an essential role in the oxidative function of
CC CC the citric acid cycle (By similarity).
CC CC -!- CATALYTIC ACTIVITY: Isocitrate + NAD(+) = 2-oxoglutarate + CO(2) +
CC CC NADH.
CC CC -!- SUBUNIT: Octamer of two nonidentical subunits IDH1 and IDH2 (By
CC CC similarity).
CC CC -!- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
CC CC -!- SIMILARITY: Belongs to the isocitrate and isopropylmalate
CC CC dehydrogenases family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; AF045153; AAC69608.1; -.
CC HSSP; P00351; IXAA.
CC DR InterPro; IPR001804; Isohd.
CC DR InterPro; IPR004434; Mito_nad_idh.
CC DR Pfam; PF00180; Isohd; 1.
CC DR TIGRfam; TIGR00175; mito_nad_idh; 1.
CC DR PROSITE; PS00470; IDH_IMDH; 1.
CC KW Oxidoreductase; NAD; Tricarboxylic acid cycle; Transit peptide;
CC FT MITOCHONDRION.
CC FT TRANSIT 1 12 MITOCHONDRION (BY SIMILARITY).
CC FT CHAIN 13 361 ISOCITRATE DEHYDROGENASE [NAD] SUBUNIT 1.
CC FT ACT_SITE 104 104 BINDING TO ISOCITRATE (BY SIMILARITY).
CC SQ SEQUENCE 361 AA; 39157 MW; 7F3D7F7C54065CAB CRC64;
CC -----
CC Query Match 48.7%; Score 38; DB 1; Length 361;
CC Best Local Similarity 63.6%; Pred. No. 23;
CC Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC -----
CC QY 2 QRLPKRMGGV 12
CC Db 20 QLLPKYGGRY 30
CC -----
CC RESULT 11
CC SX17_HUMAN
CC ID SX17_HUMAN STANDARD; PRT; 414 AA.
CC AC Q9H612;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Transcription factor SOX-17.
CC GN SOX17.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC OX NCBI_TaxID=9606;
CC RN [1]
CC RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
CC RX MEDLINE=21646335; PubMed=11786926;
CC RA Katoh M.;
CC RT "Molecular cloning and characterization of human SOX17.";
CC RL Int. J. Mol. Med. 9:153-157(2002).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
CC RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
CC RA Nakamura Y., Isegai T., Sugano S.;
CC RT "NEDO human cDNA sequencing project.";
CC RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC CC -!- FUNCTION: Probable transcriptional activator in the premeiotic
CC CC germ cells. It binds to the sequences 5'-AACAAAT-3' or 5'-AACAAAG-
CC CC 3' (By similarity).
CC CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC CC -!- TISSUE SPECIFICITY: Expressed in adult heart, lung, spleen,
CC CC testis, ovary, placenta, fetal lung, and kidney. In normal
CC CC gastrointestinal tract, it is preferentially expressed in
CC CC esophagus, stomach and small intestine than in colon and rectum.
CC CC -!- SIMILARITY: Contains 1 HMG box domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB073988; BAB83867.1; -.
CC EMBL; AK025905; BAB15277.1; -.
CC HSSP; P48436; ISX9.
CC Genew; HGNC:18122; SOX17.
CC DR InterPro; IPR000910; HMG_12_box.
CC DR
```

DR Pfam: PF00505; HMG box; 1.  
DR SMART: SM00398; HMG; 1.  
DR PROSITE: PS0118; HMG\_BOX 2; 1.  
KW DNA-binding; Nuclear protein; Transcription regulation; Activator.  
FT DNA BIND 68 136 HMG\_BOX.  
FT DOMAIN 312 351 GLN/PRO-RICH.  
SQ SEQUENCE 414 AA; 44116 MW; C78D1F24BA00ECD1 CRC64;  
Query Match 48.7%; Score 38; DB 1; Length 414;  
Best Local Similarity 70.0%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 4 LPRKMGSSYR 13  
DB 195 LPPHMGHVR 204  
RESULT 12  
GKP2\_HUMAN STANDARD; PRT; 553 AA.  
AC Q14410; 2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Glycerol kinase, testis specific 2 (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK).  
DE GN GK2 OR GKTA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=95078834; PubMed=7987308;  
RX Sargent C.A., Young C., Marsh S., Ferguson-Smith M.A., Affara N.A.; "The glycerol kinase gene family: structure of the Xp gene, and related intronless retroposons."; Hum. Mol. Genet. 3:1317-1324 (1994).  
CC -1- FUNCTION: Key enzyme in the regulation of glycerol uptake and metabolism.  
CC -1- CATALYTIC ACTIVITY: ATP + glycerol = ADP + glycerol 3-phosphate.  
CC -1- PATHWAY: Glycerol utilization; rate-limiting step.  
CC -1- SUBCELLULAR LOCATION: BOUND TO THE MITOCHONDRIAL SURFACE OR CYTOPLASMIC. IN SPERM, THE MAJORITY OF THE ENZYME IS BOUND TO MITOCHONDRIA (BY SIMILARITY).  
CC -1- SIMILARITY: Belongs to the fucokinase / gluconokinase / glycerokinase / xylulokinase family.  
CC -1- CAUTION: This could be the product of a pseudogene.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL: X78712; CAAS5365.1; -.  
DR PIR: I37417; I37417.  
DR HSP: P08859; IGLJ.  
DR Gnew; HGNC:4291; GK2.  
DR MIM; 600148; -.  
DR GO; GO:0005737; C:cytoplasm; NAS.  
DR GO; GO:0005741; C:mitochondrial outer membrane; NAS.  
DR GO; GO:0004370; F:glycerol kinase activity; NAS.  
DR GO; GO:0006071; P:glycerol metabolism; NAS.  
DR InterPro; IPR00577; FGGY\_kin.  
DR InterPro; IPR00599; Glycerol\_kin.  
DR Pfam; PF00370; FGGY; 1.  
DR Pfam; PF02782; FGGY; 1.  
DR TIGRFAMs; TIGR01311; glycerol\_kin; 1.  
DR PROSITE; PS00445; FGGY\_KINASES\_2; 1.

DR PROSITE: PS00933; FGGY\_KINASES\_1; 1.  
KW Glycerol metabolism; Transferase; Kinase; ATP-binding.  
FT NP\_BIND 167 179 ATP (PROBABLE).  
SQ SEQUENCE 553 AA; 60609 MW; 8CF53B1686BC4AD6 CRC64;  
Query Match 48.7%; Score 38; DB 1; Length 553;  
Best Local Similarity 53.8%; Pred. No. 35;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 2 QRLPKRMGSSYR 14  
DB 346 ERLAKEVGTSGYC 358  
RESULT 13  
SCAA\_CHICK STANDARD; PRT; 637 AA.  
ID SCAA\_CHICK STANDARD; Q98941;  
AC Q92075; P70095; Q98941;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Amiloride-sensitive sodium channel alpha-subunit (Epithelial Na+ channel alpha subunit) (Alpha ENaC) (Nonvoltage-gated sodium channel 1 alpha subunit) (SCNEA) (Alpha NaCH).  
DE GN SCN1A OR ENAC.  
DE OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).  
RC STRAIN=Isa brown; TISSUE=Cochlea;  
RX MEDLINE=97157073; PubMed=9003454;  
RA Killick R., Richardson G.; "Isolation of chicken alpha ENaC splice variants from a cochlear cDNA library."; Biochim. Biophys. Acta 1350:33-37 (1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Intestine;  
RA Goldstein O., Asher C., Garty H.; Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Sodium permeable non-voltage-sensitive ion channel inhibited by the diuretic amiloride. Mediates the electrodiffusion through the apical membrane of epithelial cells. Controls the reabsorption of sodium in kidney, colon, lung and sweat glands.  
CC Also plays a role in taste perception.  
CC -1- SUBUNIT: Heterotrimer of two alpha, one beta and one gamma subunit. A delta subunit can replace the alpha subunit (By similarity).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=Long;  
CC IsoId=Q92075-1; Sequence=Displayed;  
CC Name=Short;  
CC IsoId=Q92075-2; Sequence=VSP\_006196, VSP\_006197;  
CC -1- TISSUE SPECIFICITY: The long isoform has been found in cochlea, colon, and cartilage. The short isoform is only found in cochlea.  
CC -1- PTM: Ubiquitinated; this targets individual subunits for proteasome-mediated degradation (By similarity).  
CC -1- SIMILARITY: Belongs to the amiloride-sensitive sodium channel family.  
CC -1- CAUTION: It is uncertain whether Met-1 or Met-15 is the initiator.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```
or send an email to license@iab-sib.ch).
CC -----
DR EMBL: U62902; AAB50550.1; -.
DR EMBL: U62903; AAB50551.1; -.
DR EMBL: U62904; AAB50552.1; -.
DR EMBL: U58475; AAB04954.1; ALT_INIT.
DR InterPro: IPR004724; BnaC.
DR PRINTS: PR01078; AMINACHANNEL.
DR TIGRFAMs: TIGR00859; ENaC; 1.
DR Pfam: PF00858; ASC; 1.
DR PROSITE: PS01206; ASC; 1.
DR Ion transport; Sodium transport; Ionic channel; Transmembrane;
KW Glycoprotein; Ubl conjugation; Alternative splicing; Sodium channel.
FT DOMAIN 1 80
FT TRANSMEM 81 97
FT DOMAIN 98 554
FT TRANSMEM 555 571
FT DOMAIN 572 637
FT CARBOHYD 157 157
FT CARBOHYD 280 280
FT CARBOHYD 298 298
FT CARBOHYD 499 499
FT VARSPLIC 402 434
FT
FT
FT VARSPLIC 435 637
FT Missing (in isoform Short).
FT /FTid=VSP_006196.
FT Missing (in isoform Short).
FT /FTid=VSP_006197.
FT DV -> EU (IN REF. 2).
FT Q -> R (IN REF. 2).
FT SEQUENCE 637 AA; 73236 MW; 5D15E616373971DA CRC64;
Query Match 48.78; Score 38; DB 1; Length 637;
Best Local Similarity 50.08; Pred. No. 41;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Oy 1 RORLPKRMGGSY 12
Db 366 RXEWTERLGGSY 377
RESULT 14
PROS DROME
ID AC P29617; Q95GP0; Q9UGA2; Q9VGP8; PRT; 1403 AA.
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein prospero.
GN PROS OR CGI7228.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92069760; PubMed=1720353;
RA Vasselin H., Grell E., Wolff E., Bier E., Jan L.Y., Jan Y.N.;
RT "Prospero is expressed in neuronal precursors and encodes a nuclear
RT protein that is involved in the control of axonal outgrowth in
RT Drosophila.";
RL Cell 67:941-953 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92171948; PubMed=1540176;
RA Matsuzaki F., Koizumi K., Hana C., Yoshioka T., Nabeshima Y.;
RT "Cloning of the Drosophila prospero gene and its expression in
RT ganglion mother cells.";
RL Biochem. Biophys. Res. Commun. 182:1326-1332 (1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93083413; PubMed=18423358;
RA Chu-Lagraff O., Wright D.M., McNeil L.K., Doe C.Q.;
RT "The prospero gene encodes a divergent homeodomain protein that
RT controls neuronal identity in Drosophila.";
RL Development Suppl. 2:79-85 (1991).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM C).
RX MEDLINE=20503846; PubMed=11051550;
RA Xu C., Kauffmann R.C., Zhang J., Kladny S., Carthew R.W.;
RT "Overlapping activators and repressors delimit transcriptional
RT response to receptor tyrosine kinase signals in the Drosophila eye.";
RL Cell 103:87-97 (2000).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Fosler K., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclele J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RN Science 287:2185-2195 (2000).
RN [6]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Battencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX STRAIN=Berkeley; TISSUE=Head;
MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Paclele J.M., Park S., Wan K.H.,
```



DR ENBL; AFI90405; AAF06660.1; -.  
DR FlyBase; FBgn028753; Dvir\pros.  
DR InterPro; IPR007738; Prox1.  
DR Pfam; PF05044; Prox1; 1.  
KW Nuclear protein; Transcription regulation; DNA-binding;  
KW Homeobox; Developmental protein.  
FT DOMAIN 4 12 POLY-ALA.  
FT DOMAIN 19 36 ASN-RICH.  
FT DOMAIN 150 172 ASN-RICH.  
FT DOMAIN 206 209 POLY-ALA.  
FT DOMAIN 237 292 GLN-RICH.  
FT DOMAIN 305 309 POLY-ASN.  
FT DOMAIN 349 381 SER-RICH.  
FT DOMAIN 443 473 ASP-RICH.  
FT DOMAIN 544 547 POLY-ALA.  
FT DOMAIN 574 1080 GLN-RICH.  
FT DOMAIN 888 906 ASN-RICH.  
FT DOMAIN 1027 1030 POLY-ALA.  
FT DOMAIN 1045 1054 POLY-GLN.  
FT DOMAIN 1057 1062 POLY-GLN.  
FT DOMAIN 1132 1189 THR-RICH.  
FT DOMAIN 1140 1145 POLY-ALA.  
FT DOMAIN 1154 1163 POLY-GLN.  
FT DOMAIN 1183 1189 POLY-THR.  
FT DOMAIN 1090 1097 NUCLEAR LOCALIZATION SIGNAL (BY SIMILARITY).  
FT DOMAIN 1330 1337 POLY-GLN.  
FT DNA BIND 1394 1456 HOMEBOX (ATYPICAL) (BY SIMILARITY).  
FT DOMAIN 1457 1556 PROSPERO-LIKE (BY SIMILARITY).  
SQ SEQUENCE 1556 AA; 171029 MW; 6FEACFEA2D73E644 CRC64;  
Query Match 48.7%; Score 38; DB 1; Length 1556;  
Best Local Similarity 77.8%; Pred. No. 1e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 3 RLPKRMGGS 11  
Db 853 RLPKRMGGA 861

Search completed: June 3, 2004, 14:26:27  
Job time : 5.025 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 14:20:33 ; Search time 17.5 Seconds

(without alignments)  
252.415 Million cell updates/sec

Title: US-09-745-008-14

Perfect score: 78

Sequence: 1 QRLPRMGGSYRC 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

- 1: sp.archaea:\*
- 2: sp.bacteria:\*
- 3: sp.fungi:\*
- 4: sp.human:\*
- 5: sp.invertebrate:\*
- 6: sp.mammal:\*
- 7: sp.mhc:\*
- 8: sp.organelle:\*
- 9: sp.phage:\*
- 10: sp.plant:\*
- 11: sp.rodent:\*
- 12: sp.virus:\*
- 13: sp.vertebrate:\*
- 14: sp.unclassified:\*
- 15: sp.virus:\*
- 16: sp.bacteriaph:\*
- 17: sp.archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	60.3	487	16 Q8D557	Q8D557 vibrio vuln
2	47	60.3	488	16 Q87FH4	Q87FH4 vibrio para
3	44	56.4	244	12 Q8JKU8	Q8JKU8 heliothis z
4	42	53.8	107	2 O68807	O68807 synchococc
5	42	53.8	339	5 Q9VUB3	Q9VUB3 drosophila
6	42	53.8	474	16 Q89Y78	Q89Y78 bradyrhizob
7	42	53.8	1169	16 Q7UZZ9	Q7UZZ9 prochloroco
8	41	52.6	83	16 Q8YVZU7	Q8YVZU7 anabaena sp
9	41	52.6	384	16 Q7UES8	Q7UES8 rhodospirell
10	41	52.6	564	16 Q9RUK9	Q9RUK9 deinococcus
11	40.5	51.9	748	10 Q8LQH6	Q8LQH6 oryza sativ
12	40	51.3	150	16 Q82U86	Q82U86 streptomyc
13	40	51.3	151	16 Q9K4D3	Q9K4D3 streptomyc
14	40	51.3	230	16 Q8X615	Q8X615 escherichia
15	40	51.3	314	16 Q9RW7	Q9RW7 deinococcus
16	40	51.3	413	10 Q94EC2	Q94EC2 oryza sativ

17	40	51.3	472	10 Q8RZS4	Q8RZS4 oryza sativ
18	40	51.3	507	2 Q8KW84	Q8KW84 rugeria sp
19	40	51.3	538	2 Q9Z4V9	Q9Z4V9 streptomyc
20	40	51.3	559	4 Q969Y0	Q969Y0 homo sapien
21	40	51.3	885	6 Q8HYV1	Q8HYV1 sus scrofa
22	40	51.3	886	6 Q8HYV2	Q8HYV2 sus scrofa
23	40	51.3	955	4 Q8NFP4	Q8NFP4 homo sapien
24	40	51.3	1556	10 Q9MB97	Q9MB97 nicotiana t
25	40	51.3	1559	10 Q49889	Q49889 lycopersico
26	40	51.3	1955	10 Q8W078	Q8W078 oryza sativ
27	39.5	50.6	619	11 Q8BWQ5	Q8BWQ5 mus musculu
28	39	50.0	61	12 Q83177	Q83177 cauliflowe
29	39	50.0	203	4 Q8N9P6	Q8N9P6 homo sapien
30	39	50.0	277	10 Q7X801	Q7X801 oryza sativ
31	39	50.0	298	17 Q50108	Q50108 pyrococcus
32	39	50.0	332	16 Q9PI09	Q9PI09 campylobact
33	39	50.0	353	6 Q7YRY3	Q7YRY3 gorilla gor
34	39	50.0	364	3 Q873R1	Q873R1 ajellomyces
35	39	50.0	372	17 Q97U79	Q97U79 sulfobus
36	39	50.0	395	16 Q8U8A2	Q8U8A2 agrobacteri
37	39	50.0	413	16 Q7V2Y8	Q7V2Y8 prochloroco
38	39	50.0	472	2 Q9L3E9	Q9L3E9 amycolatops
39	39	50.0	520	12 Q83170	Q83170 cauliflowe
40	39	50.0	520	12 Q66163	Q66163 cauliflowe
41	39	50.0	535	4 Q86VJ4	Q86VJ4 homo sapien
42	39	50.0	542	11 Q70334	Q70334 mus musculu
43	39	50.0	575	4 Q9NXQ2	Q9NXQ2 homo sapien
44	39	50.0	737	11 Q8BJ11	Q8BJ11 mus musculu
45	39	50.0	848	11 Q8BIH6	Q8BIH6 mus musculu

#### ALIGNMENTS

RESULT 1

Q8D557 ID Q8D557 PRELIMINARY; PRT; 487 AA.  
AC Q8D557;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Mannitol-1-phosphate/altronate dehydrogenase.  
GN VV21069.  
OS Vibrio vulnificus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=672;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CMCP6;  
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
RA Choy H.E.;  
RL "Complete genome sequence of Vibrio vulnificus CMCP6.";  
RT Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; A5016811; AAC07976.1; -;  
DR GO; GO:0003824; F:catalytic activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR000669; Mannitol\_dh.  
DR Pfam; PF01232; Mannitol\_dh; 1.  
DR PRINTS; PR00084; MTL0H0GNASE.  
DR PROSITE; PS00974; MANNITOL\_DHGENASE; 1.  
KW Complete proteome.  
SQ SEQUENCE 487 AA; 53408 MW; 28786B2A15172F7D CRC64;

Query Match 60.3%; Score 47; DB 16; Length 487;  
Best Local Similarity 66.7%; Pred. No. 5.9;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 QRLPRMGGSYR 13

Db 376 QKIPQRMGGSLR 387

RESULT 2  
Q87FH4 PRELIMINARY; PRT; 488 AA.  
AC Q87FH4; 2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Mannonate oxidoreductase.  
GN VPAL705  
OS Vibrio parahaemolyticus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=670;  
[1]  
PP SEQUENCE FROM N.A.  
RC STRAIN=RMD 2210633 / Serotype O3:K6;  
RX MEDLINE=22508454; PubMed=12620739;  
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,  
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;  
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism  
distinct from that of V. cholerae.";  
RL Lancet 361:743-749(2003).  
DR EMBL; AP005089; BAC3048.1; -.  
DR GO; GO:0003824; F:catalytic activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR000669; Mannitol dh.  
DR Pfam; PF01232; Mannitol dh; 1.  
DR PRINTS; PR00084; MTLIDHNGASE.  
DR PROSITE; PS00974; MANNITOL\_DGHEINASE; 1.  
KW Complete proteome.  
SQ SEQUENCE 488 AA; 53510 MW; 700ED77E6BD56F4 CRC64;  
Query Match 60.3%; Score 47; DB 16; Length 488;  
Best Local Similarity 66.7%; Pred. No. 5.9;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 2 QRLPKRMGGSYR 13  
|:|:|:|:|:|  
Db 376 QKIPQRMGGSLR 387

RESULT 3  
Q8JKU8 PRELIMINARY; PRT; 244 AA.  
AC Q8JKU8;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Orf13.  
GN ORF13.  
OS Heliothis zea virus 1.  
OC Viruses; dsDNA viruses, no RNA stage; unclassified dsDNA viruses.  
OX NCBI\_TaxID=29250;  
[1]  
PP SEQUENCE FROM N.A.  
RC MEDLINE=22174892; PubMed=12186886;  
RA Cheng C.H., Liu S.M., Chow T.Y., Hsiao Y.Y., Wang D.P., Huang J.J.,  
RA Chen H.H.;  
RT "Analysis of the Complete Genome Sequence of the H2-1 Virus Suggests  
that It Is Related to Members of the Baculoviridae.";  
RL J. Virol. 76:9024-9034(2002).  
[2]  
PP SEQUENCE FROM N.A.  
RC Cheng C.H., Liu H.M., Hsiao Y.Y., Chow T.Y., Chen H.H.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
[3]  
PP SEQUENCE FROM N.A.  
RC Cheng C.H., Hsiao Y.Y., Liu S.M., Chow T.Y., Chen H.H.;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF451898; AAN04308.1; -.  
SQ SEQUENCE 244 AA; 28641 MW; 3DA50433EE14D855 CRC64;

Query Match 56.4%; Score 44; DB 12; Length 244;  
Best Local Similarity 53.8%; Pred. No. 10;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 1 QRLPKRMGGSYR 13  
|:|:|:|:|:|  
Db 14 KERTPKRELECTYR 26  
RESULT 4  
O68807 PRELIMINARY; PRT; 107 AA.  
AC O68807;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
OS Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=32049;  
[1]  
PP SEQUENCE FROM N.A.  
RC STRAIN=PCC7002;  
RA Droog F.N.J., Tallier B.J., Stevens S.E. Jr.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF054515; AAC08034.1; -.  
KW Hypothetical protein.  
FT NON TER 107 107  
SQ SEQUENCE 107 AA; 12163 MW; A5A44929EB610569 CRC64;  
Query Match 53.8%; Score 42; DB 2; Length 107;  
Best Local Similarity 66.7%; Pred. No. 9.6;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 2 QRLPKRMGGSYR 13  
|:|:|:|:|:|  
Db 36 QRLAKELGAHYR 47  
RESULT 5  
Q9VUB3 PRELIMINARY; PRT; 339 AA.  
AC Q9VUB3;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE CG17359 protein (RE58063p).  
GN CG17359.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
[1]  
PP SEQUENCE FROM N.A.  
RC STRAIN=Berkely.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Prannkoch C., Balgwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Borkova D., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hosain D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milehina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195 (2000).  
RN [12]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,  
RA Miranda A., Mungall C.J., Nunoo C., Pacleb J., Paragas V., Park S.,  
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RA Celniker S.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003536; AF49775.1; -;  
DR EMBL; AY071545; AAL49167.1; -;  
DR HSSP; P08045; 1ZNF.  
DR FlyBase; FBgn036396; CG17359.  
DR InterPro; IPR007087; Znf.C2H2.  
DR Pfam; PF00096; zf-C2H2; 4.  
DR ProDom; PD000003; Znf.C2H2; 1.  
DR SMART; SM00355; Znf.C2H2; 4.  
DR PROSITE; P80028; ZINC\_FINGER\_C2H2\_1; 4.  
DR PROSITE; P80157; ZINC\_FINGER\_C2H2\_2; 4.  
DR Metal-binding; Zinc; Zinc-finger.  
KW Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 339 AA; 39250 MW; 9B653D457205C57A CRC64;  
Query Match 53.8%; Score 42; DB 5; Length 339;  
Best Local Similarity 70.0%; Pred. No. 32;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 5 PKRMGGSYRC 14  
Db |||:||||  
ID Q89Y78 PRELIMINARY; PRT; 474 AA.  
AC Q89Y78;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE BLR0077 protein.  
GN BLR0077  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bradyrhizobiaceae; Bradyrhizobium.  
OX NCBI\_TaxID=375;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=USDA 110;  
RX MEDLINE=22484998; PubMed=12597275;  
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,

RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,  
RA Tabata S.;  
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110.";  
RL DNA Res. 9:189-197 (2002).  
DR EMBL; AP005935; BAC45342.1; -;  
DR GO; GO:0003824; P:catalytic activity; IEA.  
DR GO; GO:0009116; P:nucleoside metabolism; IEA.  
DR InterPro; IPR000845; PNP\_UDP.  
DR Pfam; PF01048; PNP\_UDP\_1; 1.  
KW Complete proteome.  
SQ SEQUENCE 474 AA; 52343 MW; 9537C1AE90B07BB4 CRC64;  
Query Match 53.8%; Score 42; DB 16; Length 474;  
Best Local Similarity 61.5%; Pred. No. 46;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 RQRLPKRMGGSYR 13  
Db |||:||||  
ID Q89Y78 PRELIMINARY; PRT; 1169 AA.  
AC Q89Y78;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Translation initiation factor IF-2.  
GN INFB OR PWM1494.  
OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).  
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
OC Prochlorococcus.  
OX NCBI\_TaxID=59919;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2925698; PubMed=12917642;  
RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,  
RA Acland N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,  
RA Johnson Z.I., Land M., Lindell D., Post A.F., Regalla W., Shah M.,  
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
RA Webb E.A., Zinser E.R., Chisholm S.W.;  
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic  
niche differentiation.";  
RL Nature 424:1042-1047 (2003).  
DR EMBL; BX572094; CAE19953.1; -;  
KW Initiation factor; Complete proteome.  
SQ SEQUENCE 1169 AA; 127147 MW; 0705D909F5377562 CRC64;  
Query Match 53.8%; Score 42; DB 16; Length 1169;  
Best Local Similarity 61.5%; Pred. No. 1.2e+02;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 RQRLPKRMGGSYR 13  
Db |||:||||  
ID Q89Y78 PRELIMINARY; PRT; 83 AA.  
AC Q89Y78;  
DT 01-NAR-2002 (TrEMBLrel. 20, Created)  
DT 01-NAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein Asr0358.  
GN ASR0358.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.

Deinococcaceae; Deinococcus.

EMBL; AE001983; AAF10944.1; -  
PIR; H75403; H75403.  
HSSP; P21332; LUOK.

GO; GO:0005975; p:carbohydrate metabolism  
InterPro; IPR006047; Alpha\_amyl\_cat.  
InterPro; IPR006589; A1p\_amyl\_cat sub

Pfam; PF00128; alpha-amylase  
 SMART; SM00642; Amy; 1.  
 Hydrolase: Complete proteome

Hydrolase; Complete proteome.  
SEQUENCE 564 AA; 63667 MW; B8F50B9B0DFC8D51 CRC64;

every Match 52.6%; Score 41; DB 16; Length 564;

```

Very Match      52.6%; Score 41; DB 16; Length 564;
Very Match      61.5%; Pred. No. 83;
Archives        8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

Local Similarity 61.5%; Pred. NO.  
8; Conservative 1; Mismatch

1 RQRLPKRMGGSYR 13

1 RQRLPKRMGGSYR 13

487 RQEQPALVGGSYR 499

487 RQEQPALVGGSYR 499

T 11

11 11  
96

Q8LQH6	PRELIMINARY;	PRT;	7
H6			
Q8LQH6			
Q8LQH6			

Q8LQH6 PRELIMINARY; PRT; 7  
Q8LQH6;  
01-OCT-2002 (Tremblay) 22 Created)

Q&A no. 1  
01-OCT-2002 (TrEMBLrel. 22, Created)  
01-OCT-2002 (TrEMBLrel. 22, Last sequence)

01-OCT-2002 (TREMBlrel. 22, Last sequenced)  
01-JUN-2003 (TREMBlrel. 24, Last annotated)

01-JUN-2003 (TrEMBLrel. 24, Last annot  
B1099D03.22 protein.

B1099D03.22 protein.  
B1099D03.22.  
*Oryza sativa* (japonica cultivar-group)

BI0999D03.22.  
Oryza sativa (japonica cultivar-group)  
Eukaryota: Viridiplantae: Streptophyta

*Oryza sativa* (Japanese cultivar-group); Eukaryota; Viridiplantae; Streptophyta; Spermatophyta; Magnoliophyta; Liliopsi-

Spermatophyta; Magnoliophyta; Liliopsi  
Ehrhartoideae; Oryzeae; Oryza.

Ehrhartoideae; Oryzeae; Oryza.  
NCBI\_TaxID=39947;  
[1]

```
NCBI_TaxID=39947;  
[1]  
SEQUENCE FROM N A
```

[1]  
SEQUENCE FROM N.A.  
STRAIN=cv. Nipponbare;

STRAIN=cv. Nipponbare;  
Sasaki T., Matsumoto T., Yamamoto K.;

Sasaki T., Matsumoto T., Yamamoto K.;  
"Oryza sativa nipponbare (GA3) genomic

"Oryza sativa nipponbare (GA3) genomic clone: B1099D03."; Submitted (MAR-2001) to the EMBL/GenBank

clone:BT099D03.";  
Submitted (MAR-2001) to the EMBL/GenBank  
EMBL: AP003431; BAB92645.1; -.

EMBL; AP003431; BAB92645.1; -.  
Gramene; Q8LQH6; -.

Gramene; Q8LQH6; -.  
InterPro; IPR007811; RNA\_pol\_Rpc4.

InterPro; IPR007811; RNA\_pol\_Rpc4.  
InterPro; IPR007087; Znf\_C2H2.  
Pfam: PF05132; RNA\_pol\_Rpc4; 1

InterPro; IPR007087; Znf\_C2H2.  
Pfam; PF05132; RNA\_pol\_Rpc4; 1.  
Pfam; PF00096; zf-C2H2; 3.

Pfam; PF00096; zf-C2H2; 3.  
SMART; SM00355; ZnF\_C2H2; 3.

SMART; SM00355; ZnF\_C2H2; 3.  
PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1;

PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1;  
PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2;  
Metal-binding; Zinc; Zinc-finger

PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2;  
Metal-binding; Zinc; Zinc-finger.  
SEQUENCE 748 AA: 81034 MW: 140A5F8

metal-binding; zinc, zinc-finger.  
SEQUENCE 748 AA; 81034 MW; 140A5F8

very Match 51.9%; Score 40.

Very Match 51.9%; Score 40.



```

QY 1 RORLPKRMGG 10
Db 207 KQRIEXRMGG 216

RESULT 15
Q9RWW7
ID Q9RWW7 PRELIMINARY; PRT; 314 AA.
AC Q9RWW7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Pilin, type IV, putative.
GN DE0548.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567286;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -|- SUBUNIT: THE PILI ARE POLAR FLEXIBLE FILAMENTS OF ABOUT 5.4
CC NANOMETERS DIAMETER AND 2.5 MICROMETERS AVERAGE LENGTH; THEY
CC CONSIST OF ONLY A SINGLE POLYPEPTIDE CHAIN ARRANGED IN A HELICAL
CC CONFIGURATION OF FIVE SUBUNITS PER TURN IN THE ASSEMBLED PILUS (BY
CC SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE N-ME-PHE PILI FAMILY.
DR EMBL: AE001913; AAF10127.1; -.
DR PIR: F75504; F75504.
DR TIGR: DR0548; -.
DR GO: GO:0009289; C:fimbria; IEA.
DR GO: GO:0007155; P:cell adhesion; IEA.
DR InterPro: IPR001082; Pilin.
DR InterPro: IPR001120; Prok_N_methyl_S.
DR Pfam: PF00114; pilin; 1.
DR PROSITE: PS00409; PROKAR_NTER_METHYL; 1.
KW Fimbria; Methylation; Complete proteome.
SQ SEQUENCE 314 AA; 31867 MW; 96619C5B05A3B86 CRC64;

Query Match 51.3%; Score 40; DB 16; Length 314;
Best Local Similarity 72.7%; Pred. No. 68;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RORLPKRMGG 11
Db 26 RQRGPRQGG 36

Search completed: June 3, 2004, 14:28:19
Job time : 18.5 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 14:21:19 ; Search time 6.125 Seconds  
(without alignments)  
219.866 Million cell updates/sec

Title: US-09-745-008-14

Perfect score: 78

Sequence: 1 RQRLPKRMGGSYRC 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: Piri:\*

2: Piri:\*

3: Piri:\*

4: Piri:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	100.0	1162	2 JH0557	exo-alpha-sialidas
2	41	52.6	83	2 AE1851	hypothetical prote
3	41	52.6	564	2 H75403	glycosyl hydrolase
4	40.5	51.9	235	2 JU0029	tumor necrosis fac
5	40	51.3	230	2 G65067	hypothetical prote
6	40	51.3	230	2 G91091	hypothetical prote
7	40	51.3	230	2 B85937	hypothetical prote
8	40	51.3	261	2 JQ0137	hypothetical 30.1k
9	40	51.3	314	2 F75504	probable pilin, ty
10	40	51.3	428	2 H86890	hypothetical prote
11	40	51.3	538	2 T51756	NAD+ synthase (glu
12	40	51.3	1559	2 T07757	probable DNA (cyto
13	39	50.0	298	2 B71013	hypothetical prote
14	39	50.0	332	2 G81395	probable ATP /GTP
15	39	50.0	372	2 C90498	carnitine dehydrat
16	39	50.0	395	2 A96215	hypothetical prote
17	39	50.0	395	2 AH3071	conserved hypotet
18	39	50.0	472	2 T50637	probable histidine
19	39	50.0	520	2 S11217	viropilin - caul
20	39	50.0	1638	2 D87749	protein unc-73b [1
21	39	50.0	1677	2 T14267	Xin protein, stage
22	39	50.0	2488	2 T42739	guanine nucleotide
23	38	48.7	210	2 S28673	hypothetical prote
24	38	48.7	356	2 F95954	probable dTDPgluco
25	38	48.7	422	2 T21820	hypothetical prote
26	38	48.7	553	2 I37417	glycerol kinase -
27	38	48.7	781	2 C69452	signal-transducing
28	38	48.7	876	2 T19246	hypothetical prote
29	38	48.7	1403	2 S24548	homeotic protein p

30 37 47.4 219 2 T47881  
31 37 47.4 220 2 E90022  
32 37 47.4 253 2 F64903  
33 37 47.4 253 2 H90891  
34 37 47.4 253 2 A85726  
35 37 47.4 296 2 T47062  
36 37 47.4 296 2 AC0231  
37 37 47.4 339 2 H86210  
38 37 47.4 341 2 I61725  
39 37 47.4 347 2 T06671  
40 37 47.4 348 2 A56247  
41 37 47.4 427 2 G02034  
42 37 47.4 444 2 G01924  
43 37 47.4 444 2 G01925  
44 37 47.4 455 2 G01923  
45 37 47.4 497 2 T29791

## ALIGNMENTS

### RESULT 1

JH0557

exo-alpha-sialidase (EC 3.2.1.18) - Trypanosoma cruzi

N;Alternate names: neuraminidase

C;Species: Trypanosoma cruzi

C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 11-Jan-2000

C;Accession: JH0557

R;Pereira, M.E.A.; Mejia, J.S.; Ortega-Barria, E.; Matzilevich, D.; Prioli, R.P.

J. Exp. Med. 174, 179-191, 1991

A;Title: The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial neu

A;Reference number: JH0557; MUID:91277609; PMID:1711561

A;Accession: JH0557

A;Molecule type: DNA

A;Residues: 1-1162 <PER>

A;Cross-references: GB:M61732; NID:G162302; PID:G162303

A;Note: the authors translated the codon TCT for residue 45 as Cys

C;Comment: This protein plays a role in parasite-host cell interaction.

C;Superfamily: trypanostigote-specific surface antigen

C;Keywords: glycoprotein; glycosidase; hydrolase

F;394/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 78; DB 2; Length 1162;  
Best Local Similarity 100.0%; Pred. No. 2.3e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RQRLPKRMGGSYRC 14

|||||

Db 379 RQRLPKRMGGSYRC 392

### RESULT 2

AE1851

hypothetical protein asr0358 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C;Accession: AE1851

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AE1851

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-83 <KUR>

A;Cross-references: GB:BA000019; PIDN:BA072316.1; PID:G17129703; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: asr0358

Query Match 52.6%; Score 41; DB 2; Length 83;





```
RESULT 7
B85937
hypothetical protein yqeh [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85937
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85937
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-230 <STO>
A:Cross-references: GB:AE005174; NID:g12517338; PIDN:AAG57958.1; GSPDB:GN00145; UMGF:Z41
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yqeh

Query Match 51.3%; Score 40; DB 2; Length 230;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RORLPKRMGG 10
|||:|||||
Db 207 KQRIKRMGG 216

RESULT 8
JQ0137
hypothetical 30.1K protein - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 07-Jun-1996
C:Accession: JQ0137
R:Kato, J.; Chu, L.; Kitano, K.; DeVault, J.D.; Kimbara, K.; Chakrabarty, A.M.; Misra, T
Gene 84, 31-38, 1989
A:Title: Nucleotide sequence of a regulatory region controlling alginate synthesis in P
A:Reference number: JQ0132; MUID:90108714; PMID:2514124
A:Accession: JQ0137
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-261 <KAT>
A:Note: 3-Met could also be the initiator
C:Genetics:
A:Start codon: GTG

Query Match 51.3%; Score 40; DB 2; Length 261;
Best Local Similarity 57.1%; Pred. No. 24;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RORLPKRMGSGYRC 14
|||:|||||
Db 31 RORLRKRMGSGRRRC 44

RESULT 9
F75504
probable pilin, type IV - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75504
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75504
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-314 <WHI>
```

```
A:Cross-references: GB:AE001913; GB:AE000513; NID:g6458240; PIDN:AAF10127.1; PID:g64582.
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0548
A:Map position: 1

Query Match 51.3%; Score 40; DB 2; Length 314;
Best Local Similarity 72.7%; Pred. No. 29;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RORLPKRMGGS 11
|||:|||||
Db 26 RORGPQRQGS 36

RESULT 10
H86890
hypothetical protein yvjB [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: H86890
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:111337471
A:Accession: H86890
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <STO>
A:Cross-references: GB:AE005176; PID:g12725185; PIDN:AAK06226.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yvjB

Query Match 51.3%; Score 40; DB 2; Length 428;
Best Local Similarity 77.8%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 PKRMGGSYR 13
|||:|||||
Db 280 PKKMDGSYR 288

RESULT 11
TS1756
NAD+ synthase (glutamine-hydrolyzing) (EC 6.3.5.1) [imported] - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 03-Nov-2000
C:Accession: TS1756
R:Pink, D.; Falke, D.; Wohlleben, W.; Engels, A.
Microbiology 145, 2313-2322, 1999
A:Title: Nitrogen metabolism in Streptomyces coelicolor A3(2): modification of glutamine
A:Reference number: Z25448
A:Accession: TS1756
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-538 <PIN>
A:Cross-references: EMBL:Y17736; PIDN:CAB38325.1
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: nadB
C:Superfamily: Rhodobacter capsulatus NH3-dependent NAD+ synthase
C:Keywords: ligase

Query Match 51.3%; Score 40; DB 2; Length 538;
Best Local Similarity 70.0%; Pred. No. 48;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 PKRMGGSYRC 14
|||:|||||
Db 233 PGRLGSSRC 242
```

## RESULT 12

T07757  
probable DNA (cytosine-5-)-methyltransferase (EC 2.1.1.37) - tomato  
C;Species: Lycopersicon esculentum (tomato)  
C;Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 05-May-2000  
C;Accession: T07757  
R;Cella, R.  
submitted to the EMBL Data Library, October 1997  
A;Reference number: Z16119  
A;Accession: T07757  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-1559 <CEL>  
A;Cross-references: EMBL:AJ002140; NID:e1250983; PIDN:CAA05207.1; PID:e1250984  
A;Experimental source: cultivar Meraviglia; apical meristems  
C;Genetics:  
A;Gene: SMT  
C;Keywords: methyltransferase; S-adenosylmethionine

Query Match 51.3%; Score 40; DB 2; Length 1559;  
Best Local Similarity 58.3%; Pred. No. 1.3e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRLPKRMGGSYR 13  
DB 1423 QRIKRPFGDWR 1434

## RESULT 13

B71013  
hypothetical protein PH1400 - Pyrococcus horikoshii  
C;Species: Pyrococcus horikoshii  
C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Nov-1999  
C;Accession: B71013  
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
A;Reference number: A71000; MUID:98344137; PMID:9679194  
A;Accession: B71013  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-298 <KAW>  
A;Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30506.1; PID:d1031449; PID:g32578  
A;Experimental source: strain Or3  
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C;Genetics:  
A;Gene: PH1400

Query Match 50.0%; Score 39; DB 2; Length 298;  
Best Local Similarity 63.6%; Pred. No. 41;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 RLPKRMGGSYR 13  
DB 51 RLKRLGGIYK 61

## RESULT 14

G81395  
probable ATP /GTP binding protein Cj0500 [imported] - Campylobacter jejuni (strain NCTC  
C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C;Accession: G81395  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A;Reference number: AB1250; MUID:20150912; PMID:10688204  
A;Accession: G81395  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-332 <PAR>

A;Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB75138.1; PID:g6967977

A;Experimental source: serotype O2, strain NCTC 11168

C;Genetics:

A;Gene: Cj0500

Query Match 50.0%; Score 39; DB 2; Length 332;  
Best Local Similarity 61.5%; Pred. No. 45;  
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRLPKRMGGSYR 14  
DB 57 QALAKAMGASYIC 69

## RESULT 15

C90498  
carnitine dehydratase, probable [imported] - Sulfolobus solfataricus  
C;Species: Sulfolobus solfataricus  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 01-Mar-2002  
C;Accession: C90498  
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,  
arrett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A;Description: Sulfolobus solfataricus complete genome.  
A;Reference number: A99139

A;Accession: C90498

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-372 <KUR>

A;Cross-references: GB:AE006641; NID:g13816568; PIDN:AAK43242.1; GSPDB:GN00155

C;Genetics:

A;Gene: SSO3144

C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv3272

Query Match 50.0%; Score 39; DB 2; Length 372;  
Best Local Similarity 54.5%; Pred. No. 50;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 RLPKRMGGSYR 13  
DB 209 KVPKRMGSAHR 219

Search completed: June 3, 2004, 14:29:08

Job time : 8.45833 secs